Teaching and Teacher Education 97 (2021) 103213

Contents lists available at ScienceDirect

Teaching and Teacher Education

journal homepage: www.elsevier.com/locate/tate

Research paper

How does learners' behavior attract preservice teachers' attention during teaching?

Patricia Goldberg [a,*], Jakob Schwerter [b], Tina Seidel [c], Katharina Müller [d], Kathleen Stürmer [a]

[a] University of Tübingen, Hector Research Institute of Education Sciences and Psychology, Europastr. 6, 72072, Tübingen, Germany
[b] University of Tübingen, Statistics, Econometrics and Quantitative Methods, Sigwartstr. 18, 72076, Tübingen, Germany
[c] Technical University of Munich (TUM), School of Education, Marsstraße 20-22, 80335, München, Germany
[d] Leibniz University Hannover, Institute for Education Science, Schloßwender Str. 1, 30159, Hannover, Germany

HIGHLIGHTS

- Time series analysis offers novel insight into research on teacher-student interaction.
- Preservice teachers focus more on (inter)active learning-related student behavior.
- The effect of learning-related student behavior is stable across the instructional period.

ARTICLE INFO

Article history:
Received 6 December 2019
Received in revised form
7 September 2020
Accepted 26 September 2020
Available online 17 October 2020

Keywords:
Teacher-student interaction
Teachers' attention
Student behavior
Time series analysis
Mobile eye-tracking research
Multinomial regression

ABSTRACT

Teachers need to continuously monitor students' engagement in classrooms, but novice teachers have difficulties paying attention to individual behavioral cues in all learners. To investigate these interaction processes in more detail, we re-analyzed eye-tracking data from preservice teachers teaching simulated learners who engaged in different behaviors (Stürmer, Seidel, Müller, Häusler, & Cortina, 2017). With a new methodological approach, we synchronized the data with a continuous annotation of observable student behavior and conducted time series analysis on 3646 s of video material. Results indicate that novice teachers' attention is attracted most often when learners show (inter)active learning-related behavior.

1. Introduction

A teacher's ability to provide students with sufficient learning time, engage all students in active learning processes, and elicit their cooperation comprise crucial prerequisites for enhancing students' achievement (Emmer & Stough, 2001). To manage the classroom successfully and provide appealing learning environments, teachers must direct their attention to relevant information and continually monitor students' learning (Wolff, Jarodzka, van den Bogert, & Boshuizen, 2016). Therefore, they must detect visual cues in students' behavior that indicate how learners pay attention and how engaged they are in learning content (Goldberg et al., 2019). When teachers are able to notice and identify a lack of engagement in students, they can adapt their teaching methods accordingly to encourage their students to actively engage with the learning content. However, novice teachers in particular have difficulties overseeing and distributing their attention evenly across learners (Stürmer et al., 2017; Cortina, Miller, McKenzie, & Epstein, 2015). It is assumed that they are often guided by conspicuous cues rather than an ability to monitor the classroom adequately (Wolff et al., 2016). Whereas experienced teachers observe more and notice more subtle cues compared to inexperienced teachers (e.g., Berliner et al., 1988; Carter, Cushing, Sabers, Stein, & Berliner, 1988; Sabers, Cushing, & Berliner, 1991), novice teachers are seen as

* Corresponding author.
E-mail addresses: patricia.goldberg@uni-tuebingen.de (P. Goldberg), jakob.schwerter@uni-tuebingen.de (J. Schwerter), tina.seidel@tum.de (T. Seidel), katharina.mueller@iew.uni-hannover.de (K. Müller), kathleen.stuermer@uni-tuebingen.de (K. Stürmer).

https://doi.org/10.1016/j.tate.2020.103213

lacking the required knowledge base that guides a professional view over the classroom (Berliner, 2001).

This is in line with research findings on teachers' professional vision, a concept that describes a teacher's ability to notice and interpret relevant features of classroom events for student learning (Goodwin, 1994; Sherin, 2007; van Es & Sherin, 2002). Professional vision is viewed as an indicator of knowledge representations that aid the preparation of effective teaching action (Kersting, Givvin, Thompson, Santagata, & Stigler, 2012; Sherin, 2007). When it comes to noticing relevant features, previous findings indicate that novice teachers have difficulties identifying relevant cues for teaching and learning during classroom interactions while observing videotaped classroom situations (e.g., Santagata, Zannoni, & Stigler, 2007; van den Bogert, van Bruggen, Kostons, & Jochems, 2014; Wolff et al., 2016). However, extant research also shows that novice teachers can improve this ability as part of teacher training (e.g., Sherin & van Es, 2002; Star & Strickland, 2008). In this vein, it is assumed that the underlying professional knowledge structures develop over time (Stürmer et al., 2017). Grossman et al. (2009) point out that learning to recognize relevant elements of practice comprises a crucial part of professional development, a conclusion further supported by results from expertise research indicating that regardless of the domain in question, experts have developed attentional skills that allow them to process visual information more effectively than novices (Jarodzka, Scheiter, Gerjets, & Van Gog, 2010). However, extant research on the development of teachers' ability to notice relevant cues in the classroom while teaching (i.e., in-action) remains limited. Classroom instruction is based on teacher-student interaction processes characterized by their simultaneity, multidimensionality, and immediacy (Doyle, 1986). Investigating processes related to noticing while teaching poses additional challenges compared to assessments in which novice teachers observe videotaped classroom situations (i.e., research on-action). However, identifying the engagement-related cues in student behavior that preservice teachers are able to recognize while teaching and those they do not might further improve teacher training.

In the current study, we explore novice teachers' attentional focus during instruction and aim to uncover properties of visual cues in students' behavior on which teachers fixate. To systematically assess how behavioral cues influence novice teachers' attentional focus, comparable conditions across participants are necessary. Therefore, we based our analysis on a standardized experimental setting with videos conducted by Stürmer et al. (2017) and synchronized already-existing mobile eye-tracking data from preservice teachers with a continuous annotation of learners' behavior. Deploying continuous annotation gave us a unique opportunity to analyze teacher-student interactions during instruction and investigate what kind of behavior attracts novice teachers' attention.

2. Theoretical background

Kounin (1970) identified teachers' ability to remain aware of what is going on in the classroom (*withitness*) as associated with student work involvement. Maintaining a functional overview is necessary to provide sufficient learning time, engage all students in active learning processes, and elicit their cooperation in creating a learning environment that enables all students to engage in relevant cognitive processes (Emmer & Stough, 2001). Teachers must engage in many cognitive activities to guide their students' learning (Duffy, Miller, Parsons, & Meloth, 2009). The development of so-called curriculum scripts facilitates the recognition of meaningful patterns in the classroom, which in turn enables teachers to improve their interactions with students (Putnam, 1987). Thus, as part of their expertise development, teachers need to integrate isolated knowledge structures and learn how to notice relevant cues and indicators, such as those that point out struggling students (Lachner, Jarodzka, & Nückles, 2016; Thiede et al., 2015).

2.1. Students' behavior as cues for teachers' attentional processes

When students engage in learning-relevant activities, some aspects of their cognitive processes are likely to be observable from the outside. For example, Posner (1988) demonstrated that visual orientation toward a certain stimulus improves processing efficiency. Thus, when a teacher is explaining classroom content and a student is listening, he or she might be more likely to turn and face the person speaking in order to better process the relevant information. This kind of student behavior, which can be described as external and observable activity, is viewed as an important element of the larger, multi-dimensional construct of student engagement (Fredricks, Blumenfeld, & Paris, 2004), as well as one of the key elements of learning and academic success. Three types of engagement have been defined: cognitive, emotional, and behavioral (Fredricks et al., 2004). While psychological investment in learning (cognitive component) and affective reactions to classroom situations (emotional component) are more internal processes, the behavioral component is observable. Concentration, attention, asking questions, and contributing to class discussions are activities that are already known to signal certain learning-related processes and become observable in students' behavior to some extent (Fredricks et al., 2004). As the three components are highly interrelated and do not occur in isolation, students' overt behavior can provide visible indicators of whether they are engaged in appropriate learning-related processes, which are in turn an important determinant of academic achievement (Lahaderne, 1968; McKinney, Mason, Perkerson, & Clifford, 1975). Previous research has found correlations between students' behavioral engagement and academic achievement (Lei, Cui, & Zhou, 2018), as well as between students' attention-related behavior and achievement (Helmke & Renkl, 1992; Hommel, 2012; Karweit & Slavin, 1981; Stipek, 2002). Opposing results finding no relation to achievement (e.g., Pauli & Lipowsky, 2007) might be due to the applied survey method (self-reports vs. observer ratings) and a restricted focus on certain facets of learning-related behavior. For example, measuring only active on-task behavior (Lipowsky, Rakoczy, Pauli, Reusser, & Klieme, 2007), without considering off-task behavior, does not account for the broad behavioral spectrum that students might demonstrate during classroom instruction, and thus does not allow for detection of possible effects of other kinds of behavior.

The challenge for teachers lies in noticing behavioral cues that are relevant for inferring individual students' needs. However, interpreting student behavior is not always straightforward and depends on both students' learning activities and their individual prerequisites. Learners can differ in their learning-related behavior, but still all be engaged in a certain task. Simultaneously, a lack of certain behaviors can pinpoint a student who is distracted or whose mind is wandering. Therefore, students' learning-related behavior differs with respect to the learning activities in which they are engaged (Chi & Wylie, 2014). For example, previous research shows that high-achieving students typically engage more verbally than low-achieving students (e.g., Kelly, 2008; Sacher, 1995), and students with stronger beliefs in their own competence participate more often in classroom discussions than less-confident students (Böheim, Knogler, Kosel, & Seidel, 2020; Pauli & Lipowsky, 2007). Additionally, profiles based on students' general cognitive abilities, acquired knowledge in subject domains, interest, and subject-related self-concept (Seidel, 2006) can predict students' verbal participation (Jurik, Gröschner, & Seidel, 2013). Thus, the interplay between cognitive and motivational-affective prerequisites affects observable student behavior in teacher-student interactions. However, as previously mentioned, students' activities fall across a broad behavioral spectrum. Depending on their individual prerequisites, some students might display rather salient and active behavior, such as participation in classroom discussions or disruptions, whereas other students might remain unobtrusive and passive (Seidel, Schnitzler, Kosel, Stürmer, & Holzberger, 2020). Salient behavior is easier to observe, and teachers might have fewer difficulties inferring cognitive processes in more active students compared to quieter students with more subtle actions, even though the latter group might actually need the teacher's attention because they are struggling. Therefore, it is important that teachers not only react to salient student behavior, but also notice subtle cues that indicate problems and obstacles. Additionally, it is crucial that teachers are able not only to differentiate between attentive and non-attentive students, but also to determine the underlying cause of inattention (e.g., not interested vs. struggling; Seidel et al., 2020).

2.2. Measuring teachers' attention

To design effective teaching, teachers need to develop professional vision skills that allow them to identify important events and cues during teacher-student interactions (van Es & Sherin, 2002). However, previous research indicates that the required knowledge base is not yet present in novice teachers, but rather develops over time (Berliner, 2001). Novice teachers have been shown to have problems noticing relevant aspects of classroom instruction compared to more experienced teachers. For example, early research has demonstrated that expert teachers are better at noticing subtle differences in instructional strategies (Sabers et al., 1991) and that novices have difficulties focusing on students' actions (Carter et al., 1988). Following Blomberg, Stürmer, & Seidel (2011), *noticing* describes teachers' ability to pay attention to important aspects in complex classroom environments. To measure teachers' noticing ability, video prompts (Seidel & Stürmer, 2014; Stürmer & Seidel, 2015), questionnaires (Steffensky, Gold, Holodynski, & Möller, 2015), and/or qualitative analysis of open questions (Kersting, 2008; van Es & Sherin, 2008) are deployed. However, using such non-physical measurements only can provide limited information on teachers' attentional focus, as these processes might happen rather unconsciously. Using attentional skills as an indicator of expertise, eye-tracking technology already has been used to study professional vision in various domains. The specialized way that members of a professional group view a scene of interest has been shown to be domain-independent and connected to expertise level. Due to their well-organized and structured schemata of concepts (Chi, Glaser, & Rees, 1982), experts possess attentional skills that allow them to focus on relevant rather than irrelevant visual information (Jarodzka et al., 2010). For example, experts were shown to fixate more often on relevant rather than irrelevant areas during chess games (Charness, Reingold, Pomplun, & Stampe, 2001). When viewing dynamic stimuli, experts exhibit longer, but fewer, fixations on relevant areas, indicating that experts might exhibit more selective search strategies because they know the visual cues that provide important information (Moreno, Reina, Luis, & Sabido, 2002). As these studies indicate, experts and novices differ in how they view certain situations and how they perceive visual information. Thus, teachers' visual perception can also provide important insights into their ability to notice relevant information within complex classroom interactions (Lachner et al., 2016). However, this complexity poses additional challenges in terms of attention allocation that distinguish research on teaching from the aforementioned studies (Cortina et al., 2015).

As teaching is defined as a process of teacher-student interaction, students also influence teachers' behavior. For example, they might interact through explicit behavior, such as asking questions or disturbing classroom instruction, or subliminal behaviors, such as showing a lack of understanding through their facial expressions. In this context, distinguishing relevant from irrelevant information becomes more complex, as teachers must interact with their students and react to contextual demands simultaneously. For example, during classroom discussions, teachers need to listen to student answers, consider the relevance and quality of these answers, and think about the next question, while simultaneously scanning the class for misbehavior and/or signs of miscomprehension (Doyle, 1986). Consequently, inexperienced teachers can easily become overwhelmed because they are not yet able to process all incoming information effectively and decide which visual cues are most relevant. Due to excessive demands, processes that direct novice teachers' eye movements might differ from those of experts. Human eye movements in general are guided by two broad processes: bottom-up, through salient features in targets, and top-down, such as through plans and intentions derived from professional knowledge (Seidel et al., 2020; Schütz, Braun, & Gegenfurtner, 2011; Shulman, 1987). Therefore, it can be assumed that these processes also drive teachers' visual attention while teaching (Lachner et al., 2016). On one hand, salient features such as students raising their hands or disturbing the classroom can catch teachers' attention. On the other hand, teachers' attention also can be driven by specific tasks when observing certain students more closely, such as gathering information about their cognitive processes. This intentional distribution of attention requires more top-down mechanisms and has been shown to be associated with teaching expertise (Haataja et al., 2019; McIntyre, Mainhard, & Klassen, 2017). Psychological studies in the field of attention research further indicate that bottom-up processes initially guide visual attention, before intentional, top-down processes intervene and control the attentional focus (Theeuwes, Atchley, & Kramer, 2000). Therefore, it can be assumed that alongside expertise, a temporal component impacts how teachers' attention is guided during instruction.

2.3. How preservice teachers' attention is guided during instruction

By analyzing classroom videos, Lipowsky et al. (2007) found that teachers tend to interact with high-performing students and actively engage with them more often compared with weaker students. However, interaction with students alone does not capture the actual focus of teachers' attention. Past research has deployed eye-tracking technology to investigate teachers' ability to detect relevant events in classroom scenarios (van den Bogert et al., 2014; Wolff et al., 2016; Yamamoto & Imai-Matsumura, 2015). However, these studies' findings are limited with respect to external validity, as participants' eye movements are recorded while they look at a screen showing an instructional setting, as opposed to engaging in a real classroom with teacher-student interactions. As previous research demonstrates that people's gaze behavior in laboratory settings differs from that in the real world (Foulsham, Walker, & Kingstone, 2011), teachers might also perceive a classroom situation differently when watching it on a computer screen (on-action) compared to actually being in the situation (in-action).

Recent in-action research has deployed mobile eye-tracking technology to study teachers' cognitive load (Prieto, Sharma, Wen, & Dillenbourg, 2015) or compare teachers' gazes for information-seeking and information-giving across expertise and culture (McIntyre et al., 2017). Furthermore, Cortina et al. (2015) assessed expert and novice teachers' eye movements during teaching with mobile eye-tracking technology. Novice teachers tended to give their undivided attention to particular students while providing feedback, while expert teachers were capable of monitoring the whole classroom simultaneously. These results are supported by Dessus, Cosnefroy, and Luengo (2016), who investigated teachers' strategies with respect to expertise. Experienced teachers were able to distribute their attention more frequently to a broader set of students than novice teachers. Stürmer et al. (2017) found similar results, as preservice teachers distributed their attention unevenly across four learners with different learning prerequisites while teaching in standardized settings. Notably, preservice teachers focused their attention on their instructional material 30.24% of the time (Stürmer et al., 2017). Furthermore, when looking at learners, all preservice teachers mainly focused on one learner, even though they did not focus consistently on learners who shared the same set of individual prerequisites (Stürmer et al., 2017). Similarly, Dessus et al. (2016) assumed that teachers' gaze might depend on certain salient student characteristics, and therefore considered students' current subject performance as well as self-reported and teacher-perceived behavioral self-regulation abilities in their analysis. Their results suggest that students' level of performance and self-regulation might affect experienced teachers' gaze, but not novice teachers' gaze. Taken together, other explanations besides student characteristics might guide novice teachers' attentional focus. According to findings by Wolff et al. (2016), inexperienced teachers' attentional processes might be driven rather bottom-up through salient features in student behavior rather than their intention to diagnose students' cognitive processes (top-down; Schütz et al., 2011). However, existing research has yet to examine what has happened in the classroom by the time students capture preservice teachers' attention.

2.4. Research questions

Current approaches do not consider teacher-student interactions in more detail, and research on how student behavior affects novice teachers' attention in particular during instruction is lacking. Therefore, in the present study, we investigate these interactions for the first time in an exploratory manner by analyzing preservice teachers' attentional processes contingent upon students' behavior in a small sample of video material. Despite the rather small sample size, the videos display standardized teaching situations with comparable behavior by learners. These standardized teaching situations involved preservice teachers instructing a small group of learners in a setting with reduced complexity on the same domain-independent topics. Learners acted in accordance with profile scripts so that the circumstances were the same for all preservice teachers (Seidel, Stürmer, Schäfer, & Jahn, 2015). Thus, the videotaped settings offer a unique opportunity to uncover properties of visual cues in learners' behavior that novice teachers fixate on, and to examine the stability of these effects over the course of preservice teachers' teaching.

To control for possible confounding effects within the complexity of teaching, it is important to ensure standardized conditions. For research on-action (e.g., observing videotaped classroom situations), this implies, for example, using the same video material for all participants. However, providing similar situations in research in-action is more complicated, as much variation exists across the spectrum of students and their behavior. While Cortina et al. (2015) compared expert and novice teachers' attentional processes while instructing the same classrooms with the same students, Seidel et al. (2015) developed standardized "training" situations to provide comparable conditions for preservice teachers in their first teaching experiences. We based our analysis on Stürmer et al. (2017) video data, in which seven preservice teachers were asked to teach four simulated learners in a standardized teaching situation. The lesson topics were predefined (tactical game, public transportation system), and the instructional time lasted for a maximum of 20 min. Learners comprised university students who were carefully trained and systematically assessed to behave in accordance with either an uninterested (mixed cognitive abilities, low interest), underestimating (high cognitive abilities and prior knowledge, low self-concept, intermediate level of interest), struggling (low cognitive abilities, knowledge, and self-concept), or strong (high cognitive abilities, knowledge, self-concept, and interest; Seidel, 2006) profile. Acting scripts provided background information about each profile in terms of cognitive and motivational-affective characteristics, as well as observable behavioral indicators. The strong profile was instructed to interact with the preservice teacher in an active and motivational manner, whereas the underestimating profile would only participate actively when directly engaged and made comments indicating a lack of confidence. The uninterested profile was instructed to actively exhibit low interest and engage in disturbing behaviors and comments, while the struggling profile would exhibit avoidant, shy behavior and try not to become actively engaged in interaction with the teacher (see Seidel et al., 2015). The learners were taught to act using observable behavioral indicators and further instructed to interact naturally and adapt their behavior in line with the teaching-learning process taking place in the situation (Stürmer et al., 2017).

To identify specific interaction patterns between preservice teachers' attentional focus and what is occurring in the instructional setting, we applied a new methodological approach to the data sources in which we synchronized preservice teachers' mobile eye-tracking data with a continuous rating of visible cues in learners' behavior, ranging from salient to rather unobtrusive indicators, and conducted time series analysis. The following research questions were addressed:

1) Are there behaviors in simulated learners that capture preservice teachers' attention? Does salient behavior capture preservice teachers' attention relatively more often compared with less salient behavior?
2) Does the effect of learners' behavior on preservice teachers' attention change over time?
3) Are there profile-specific differences in how learners' behavior affects preservice teachers' attentional focus?

3. Method

3.1. Sample and procedure

To answer our research questions, we based our analysis on the data from Stürmer et al. (2017) eye-tracking study, where seven preservice teachers taught one out of two pre-defined topics in a standardized teaching situation. The seven preservice teachers constituted a subsample of a full cohort of preservice teachers ($N = 89$, age: $M = 22.2$ years, $SD = 2.0$, 56% *female)* from the teacher education program at the Technical University of Munich (TUM), Germany. The program focuses on training secondary school mathematics and science teachers. The full cohort participated in the standardized teaching situations in their third year of the teacher education program as part of a university course (see Seidel et al., 2015). At this point, the cohort had already gathered some teaching experience by successfully completing three short internships in schools and classrooms. However, as the preservice teachers were about to begin their professional teacher preparation program, they could be regarded as novices in teaching. The study by Seidel et al. (2015) investigated to what extent these novices display teaching skills in the standardized situations, identifying differences in preservice teachers' teaching quality (e.g., structuring, teaching support, and learning climate), and validated the shown teaching skills with real classroom performance. Within the sample, preservice teachers were asked to voluntarily participate in an eye-tracking study (Stürmer et al., 2017). A total of seven preservice teachers ($n = 5$ female) wore eye-tracking glasses while teaching in the standardized situations (age: $M = 22.19$ years, $SD = 2.3$). This subsample can be considered as representative for the full study cohort, as they did not deviate more than one standard deviation from the cohort means on measures of their motivational learning prerequisites (*ability self-concept with regard to teaching*: full cohort $M = 3.44$, $SD = 0.45$/subsample $M = 3.67$, $SD = 0.32$, scale from 1 = does not apply to 4 = applies; *self-efficacy with regard to teaching:* full cohort $M = 2.96$, $SD = 0.32$/subsample $M = 3.23$, $SD = 0.39$, scale from 1 = does not apply to 4 = applies), the way they adapted to the teaching role in the situation (full cohort $M = 3.80$, $SD = 0.35$/subsample $M = 3.94$, $SD = 0.10$, external rating from 1 = does not apply to 4 = applies) and with regard to their shown teaching skills (*structuring:* full cohort $M = 1.67$, $SD = 0.45$/subsample $M = 1.46$, $SD = 0.35$; *teaching* support*:* full cohort $M = 1.92$, $SD = 0.66$/subsample $M = 1.67$, $SD = 0.52$; *learning climate:* full cohort $M = 2.46$, $SD = 0.41$/subsample $M = 1.42$, $SD = 0.13$, external ratings from 1 = does not apply to 4 = applies). For our data analysis, we had to reduce the sample size from the pool of seven videos, as two of the original eye-tracking datasets could not be synchronized. Furthermore, one preservice teacher's instructional time in the standardized setting was much shorter; thus, the range of behavior learners were supposed to provide was not comparable. The four videotaped sessions totaling $N = 3646$ s on which our analysis is based are comparable in length (Table 1) and in the ways the simulated learners acted (see Fig. 1). In the original data, each session was video-recorded with a complete view of the situation, and preservice teachers wore mobile eye-tracking glasses. Preservice teachers and simulated learners were placed around two tables, with the underestimating and uninterested learners sitting on the right-hand side of the preservice teacher and the strong and struggling learners on the left-hand side. The seating order was kept constant across participants. Each of the four simulated learners was defined as one area of interest (AOI).

3.2. Analysis

Behavior annotation. In the current study, we manually annotated learners' observable behavior on a one-dimensional scale over the entire instructional period in 1-s steps. The free software CARMA (Girard, 2014) enables continuous interpersonal behavior annotation using joysticks (see Lizdek, Sadler, Woody, Ethier, & Malet, 2012). We combined the idea of on-task/off-task behavior (Helmke & Renkl, 1992; Hommel, 2012) with existing scales from the engagement literature and used the ICAP framework (Chi & Wylie, 2014) as inspiration to define more fine-grained differentiations within the spectrum of possible behaviors (passive, active, (de)constructive, and interactive). Thus, behavior was annotated on a symmetric scale ranging from −2, which indicated disruptive (i.e., interactive) off-task behavior, such as shouting across or walking around the classroom with the intention to interrupt, to +2, indicating highly engaged, interactive, on-task behavior in which, for example, learners ask questions and try to explain content to fellow learners (see Fig. 2). Values closer to 0 indicated rather unobtrusive, passive behavior in which, for example, learners listened without participating (on-task) or rummaged through their belongings (off-task; Goldberg et al., 2019). Two raters annotated each learner in all videos in random order, with inter-rater reliability ICC(2,1) for each student profile ranging between 0.75 and 0.83 (absolute agreement). For the subsequent analysis, the mean of the two raters was calculated for every learner at every second. In addition to the effect of behavior in general, we also investigated the impact of especially salient (i.e., active and interactive) behavior. To account for different effects of salient on-task and salient off-task behaviors, we defined behavioral annotation values above 1 and below −1 as salient behaviors and calculated two binary variables.

Teacher event rating. To enrich our analysis and control for the specific instructional setting, we conducted an event rating of preservice teachers' instructional practices (e.g., asking questions). Two raters coded the events according to the category system displayed in Fig. 3, with an inter-rater reliability of $k = 0.64$ (good agreement; Döring & Bortz, 2016). The raters applied a binary overall classification of preservice teachers' behavior (talks or does not talk) and also indicated whenever one of the students addressed the preservice teacher.

Preparing the time series. To reduce the information from the eye-tracking data and synchronize it with the manual annotations, we only used the fixation with the longest duration for each second. As a result, we conducted a time sampling of preservice teachers' AOIs and each simulated learner's behavioral information on a per-second basis. The resulting dataset is a time series that specifies preservice teachers' AOIs, the behavioral score for each learner, whether learners showed salient on- or off-task behavior, and what the preservice teacher did (i.e., teacher events) for each second.

Statistical analysis. We wanted to predict preservice teachers' AOIs, that is, whether they fixated on the underestimating, uninterested, strong, or struggling learner. Preservice teachers' AOIs are by nature a multinomial variable, as we cannot order the different profiles into a hierarchy of better or worse. Therefore, we applied multinomial regressions by using a mixed model with alternative-specific and alternative-unspecific variables. We predicted preservice teachers' AOIs based on learners' behavioral ratings with a time lag. This decision was made to overcome the question: Which happens first, if both measures − AOI and behavioral rating − are used for the same second? By using time lags, we allow for a causal interpretation of the findings, as preservice teachers' AOIs should not influence learners' behavioral ratings one or more seconds earlier; that is, reverse causality is not an issue. We used the behavioral rating (first time lag; subsequently referred to as rating) together with the variables indicating salient on- and off-task behavior 1 s before preservice teachers' AOIs as our main variables of interest and also included the second and third time lag of the ratings to control for autocorrelation in our regressors.

Aside from the four students, the preservice teachers could choose not to look at any of the four learners, but rather somewhere in the room or at their instructional material. We used this option as the alternative in the multinomial regression, giving gaze towards the room/instructional material a rating score of zero. All variables of interest are alternative-specific, which eases the

Table 1
Video length for each preservice teacher.

Preservice teacher	Total seconds	Total minutes
1	1044	17.40
2	971	16.18
3	988	16.47
4	888	14.80

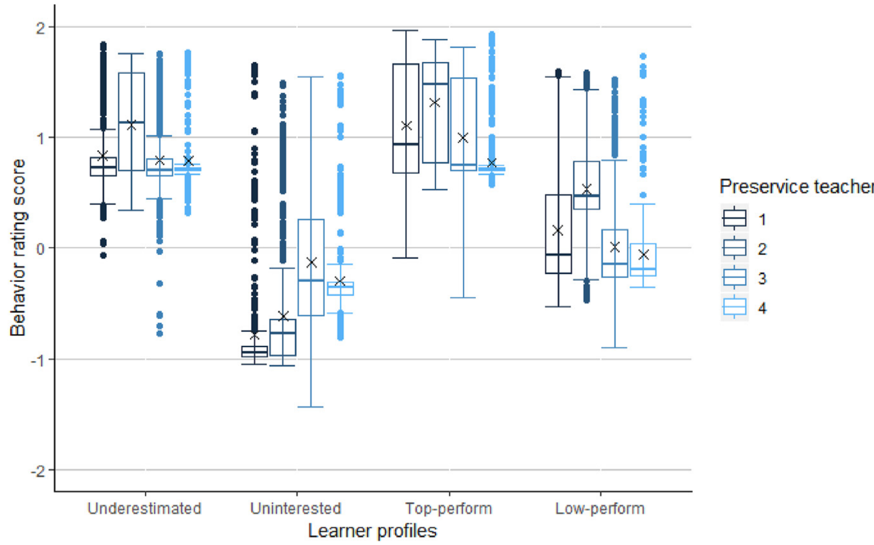

Fig. 1. Behavior ratings per learner profile as boxplots separately for each preservice teacher (x indicating the mean).

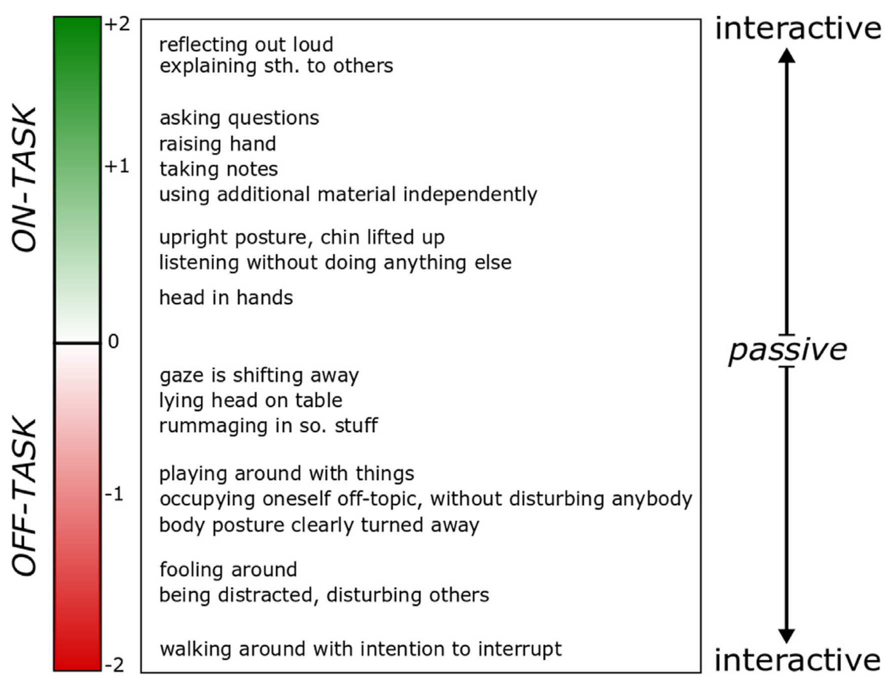

Fig. 2. Scale for behaviour annotation with example behavioural indicators. From Goldberg et al. (2019). CC BY.

interpretation of the regression. This means that because we have individual ratings (and salient on- and off-task behavior) for all learners, we get one coefficient for the rating (and salient on- and off-task behavior respectively) for all alternatives (Cameron & Trivedi, 2005). By comparison, the teacher-event variables are not alternative-specific, but are the same for all learners. This leads to individual coefficients for each teacher event for each of the four alternatives.

To check for profile-specific effects, we conducted linear probability models for each learner profile separately. For this, we recoded the multinomial outcome variable as a series of binary variables, that is, a series of dummy variables equal to one if, for example, the preservice teachers' AOIs were directed towards the underestimating learner and zero otherwise. Regressions were calculated for the uninterested, strong, and struggling learners respectively, as well as for the alternative in which the preservice teachers looked anywhere but at one of the learners. We included all four preservice teachers in the analysis and controlled for general differences between the teachers by including dummy variables for each teacher. These analyses include a coefficient for salient off-task behavior only for the uninterested learner, as the other learners displayed no such behavior.

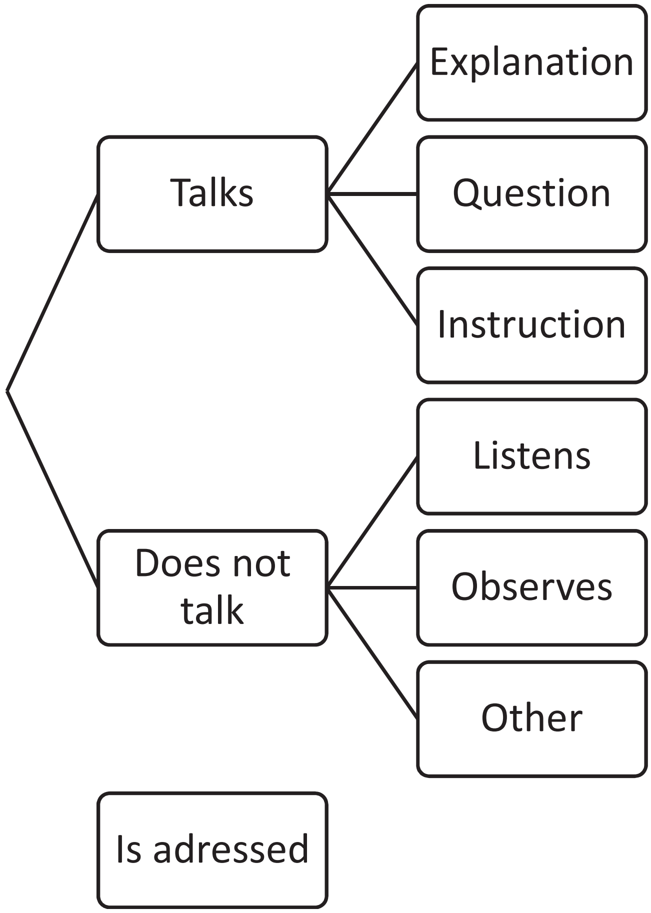

Fig. 3. Event coding system of preservice teacher behavior.

4. Results

4.1. Influence of learners' behavior on preservice teachers' attentional focus

In a first step, we included all preservice teachers in one multinomial regression. The coefficient of the manual annotation of learning-related behavior was significantly positive, $b = 2.04$, $p < .001$, which means that the more learners' behavior moves toward the interactive on-task end of the behavioral continuum, the higher the likelihood that a learner will be looked at by the preservice teacher in the next second. Inversely, the more learners' behavior moves towards the interactive off-task end of the behavioral continuum, the lower the likelihood that a learner will be looked at by the preservice teacher in the next second. There was also a significant positive relationship with whether or not a learner showed salient on-task behavior, $b = 0.24$, $p < .05$. Thus, engaging in behavior such as asking questions or explaining something increased the probability of being in the preservice teachers' AOIs. Whether or not a learner displayed salient off-task behavior showed no significant relationship, $b = -0.85$, $p = .052$. Regarding gaze stability, the behavioral rating 2 s earlier was significantly negative, $b = -1.36$, $p < .001$, while the behavioral rating 3 s earlier was not significant, $b = 0.27$, $p = .263$. This means that if learners' behavior had a high rating score, the preservice teachers were less likely to keep looking at the respective learner 2 s later. The behavioral rating 3 s earlier did not exert any effect. To show that these effects are not sensitive to our choice of specification, Table 2 depicts the results when not controlling for the linear time trend (Model 2), preservice teachers (Model 3), teacher events (Model 4), or all of these (Model 5).

In multinomial regressions, only the direction and significance of the alternative-specific coefficient can be interpreted directly; the numerical value of the coefficient itself cannot because of the multinomial model's non-linearity. Therefore, we also calculated the marginal effects at the mean for the rating and for salient on-task behavior (Table 3). Values on the diagonal indicate the percentage increase in the likelihood of being looked at by the preservice teacher if the rating score rises by one unit or salient on-task behavior is shown. For example, if all variables are equal to their means, and the underestimating learner's rating score increases by one unit, the probability that the preservice teacher fixates on this learner increases by 18.09%. Values off the diagonal, in turn, indicate the percentage with which the likelihood of being in the teacher's AOI decreases when another learner's rating score rises, or if this other learner shows salient on-task behavior. The effect is symmetric, that is, an increase in the rating score of the underestimating learner, for example, leads to an equal decrease in the probability of the uninterested learner being in the preservice teacher's AOI (by 3.52%), as an increase in the rating score for the uninterested learner decreases the probability of the preservice teacher fixated on the underestimated learner. This is a general feature of alternative-specific regressions.

To see whether our results are driven by just one preservice teacher and cannot be generalized, we ran the multinomial regressions separately for each preservice teacher. We found the same underlying patterns as in the aforementioned regression results (in which we included all teachers), with only minor deviations: When analyzing each preservice teacher separately, we again found a positive effect of the rating and a negative effect of the rating 2 s earlier. Therefore, we conclude that the effect of learners' behavior is not specific to one preservice teacher and thus is more generally valid. However, for salient on- and off-task behavior, we found mixed results in the teacher-specific regressions (for more information on the exact regression analysis, see Figs. A1 and A3 as well as Tables A2 and A4 in Appendix A).

4.2. Influence of time on preservice teachers' attentional focus

Next, we investigated the impact of elapsed time during the course of instruction. Starting at the time point of 80 s[1], we calculated regressions by adding data from the next second and continued the calculations over the time course. Fig. 4 shows how the different coefficients help explain the teachers' AOIs over the course of instruction. When the full 95% confidence interval (as indicated by the blue area in the figure) is above or below zero, the coefficient's effect is significant at the 95% significance level.

The rating coefficient shows a stable positive effect, which increases only marginally after 500 s. As no changes exist over time and the rating's effect does not depend on the time point within the instructional period, the effect of the rating can be viewed as stable over time. By comparison, the coefficient for salient on-task behavior shows some instability at the beginning of the instructional period, but this effect also stabilizes after about 500 s and appears to be similarly robust thereafter. This is not the case for the second time-lag coefficient of the rating. The estimation is less precise, and the effect is significant only after the teacher spent 700 s with the learners. Additionally, the estimation of salient off-task behavior is the most imprecise, as the confidence interval for

[1] We started with the 80th second to have a sufficient number of observations and enough variation in the data to calculate reasonable results.

Table 2
Prediction of preservice teachers' AOI (N = 3618 s).

	Model 1			Model 2			Model 3			Model 4			Model 5		
	b	SE	p	b	SE	p	b	SE	p	b	SE	p	b	SE	p
Rating $_{t\text{-}1}$	2.04	0.25	<.001	2.05	0.25	<.001	2.31	0.25	<.001	2.08	0.25	<.0001	2.34	0.25	<.001
Salient on-task behavior $_{t\text{-}1}$	0.24	0.10	.019	0.26	0.10	.009	−0.11	0.091	.023	0.48	0.10	<.001	0.24	0.08	.005
Salient off-task behavior $_{t\text{-}1}$	−0.85	0.44	.052	−0.915	0.43	.035	−0.76	0.43	.075	−0.68	0.43	.114	−0.63	0.42	.138
Rating $_{t\text{-}2}$	−1.36	0.41	.001	−1.39	0.40	.001	−1.45	0.41	<.001	−1.48	0.40	<.001	−1.63	0.40	<.001
Rating $_{t\text{-}3}$	0.27	0.24	.263	0.263	0.24	.135	0.34	0.24	.151	0.31	0.24	.194	0.48	0.24	.041
Controlled for teacher events		Yes			Yes			Yes			No			No	
Controlled for teacher		Yes			Yes			No			Yes			No	
Controlled for linear time trend		Yes			No			Yes			Yes			No	
x^2		1790*			1719*			1475*			1422*			1019*	
Pseudo R^2		.157			.151			.129			.125			.089	

Note: x^2 refers to the Likelihood Ratio Test. We calculated McFadden's Pseudo R^2. *p < .001.

Table 3
Marginal effects at the mean for the coefficients in percent.

	Profile	underestimating	uninterested	strong	struggling
Rating	underestimating	23.62%	−4.73%	−6.56%	−3.86%
	uninterested		29.20%	−8.50%	−5.01%
	strong			37.19%	−6.94%
	struggling				24.76%
Salient on-task	underestimating	2.73%	−0.05%	−0.08%	−0.45%
	uninterested		3.38%	−098%	−0.58%
	strong			4.30%	−0.80%
	struggling				2.86%

Note: As the matrix is symmetrical, only the upper part is reported here.

this coefficient is the widest. The development over time here is also the least stable, given the different ups and downs of the coefficient. Thus, the effects of salient off-task behavior cannot be viewed as stable across instructional time.

4.3. Profile-specific effects of learners' behavior

Finally, we calculated separate linear probability models for each student profile to investigate profile-specific effects of learners' behavior. The binary regression results are summarized in Fig. 5, showing the variables of interest with their 95% confidence intervals (for exact values, see Appendix B).

We found a significant effect of the rating for all profiles: If the rating increased by one unit, the probability of the preservice teacher focusing on that specific learner also increased. Thus, for example, asking questions or explaining something increased the probability of being in the preservice teachers' AOIs for all profiles. However, the rating exerted the greatest impact on the strong profile and the weakest impact on the struggling profile, suggesting the presence of profile-specific effects. Additionally, for some profiles, the other profiles' behavioral rating exerted a significant effect on the probability of being in the preservice teacher's AOI: When the uninterested and strong learners' ratings increased, the probability that the preservice teacher would fixate on the underestimating learner decreased. Similarly, a rise in the underestimating and uninterested learners' behavioral ratings decreased the probability of the strong learner being in the preservice teachers' AOI, while an increase in struggling learner's rating decreased the probability of the uninterested learner being in the preservice teachers' AOIs. An increased behavioral rating for the other learners did not significantly affect the probability of the preservice teacher fixating on the struggling learner. The ratings 2 s earlier only exerted a significant effect on the strong learner. Similar to the results for the multinomial regression, a high rating score for the strong learner 2 s earlier decreased the probability that the preservice teacher would keep looking at him or her.

To cross-check our results, we also ran the linear probability model for the alternative case (i.e., the preservice teacher not looking at any learner). As expected, the learners' behavioral ratings did not explain when the preservice teacher looked elsewhere in the room or at the instructional material.

5. Discussion

In the present study, we aimed to more closely investigate student-teacher interactions by synchronizing preservice teachers' eye-tracking data with a continuous annotation of learners' behavior. We used time series analysis to examine whether certain behaviors in learners provoke preservice teachers' attentional focus and what role salient behaviors play in particular. Additionally, we evaluated the impact of the time point within instruction on novice teachers' attention. As students with different individual characteristics exhibit different kinds of visual cues, we further investigated profile-specific effects of learners' behavior on preservice teachers' attentional focus.

The patterns found in our results support previous research on preservice teachers' monitoring skills. Like Lipowsky et al. (2007), we found that preservice teachers focus their attention on students who are engaging in more (inter)active learning-related behavior, especially salient on-task behavior. Thus, active participation, such as asking questions or explaining something, increased the likelihood of the preservice teacher focusing on a learner who displayed this kind of behavior. Furthermore, the less actively learners participate and the more distracted their behavior becomes, the lower the probability of the preservice teacher focusing his or her attention on them. A possible explanation may concern teachers' need to control instructional progress (see Hofer, 1997). Novice teachers in particular might be more sensitive to this desire for control compared to experienced teachers. We found that preservice teachers focused more on learners who showed behavior that

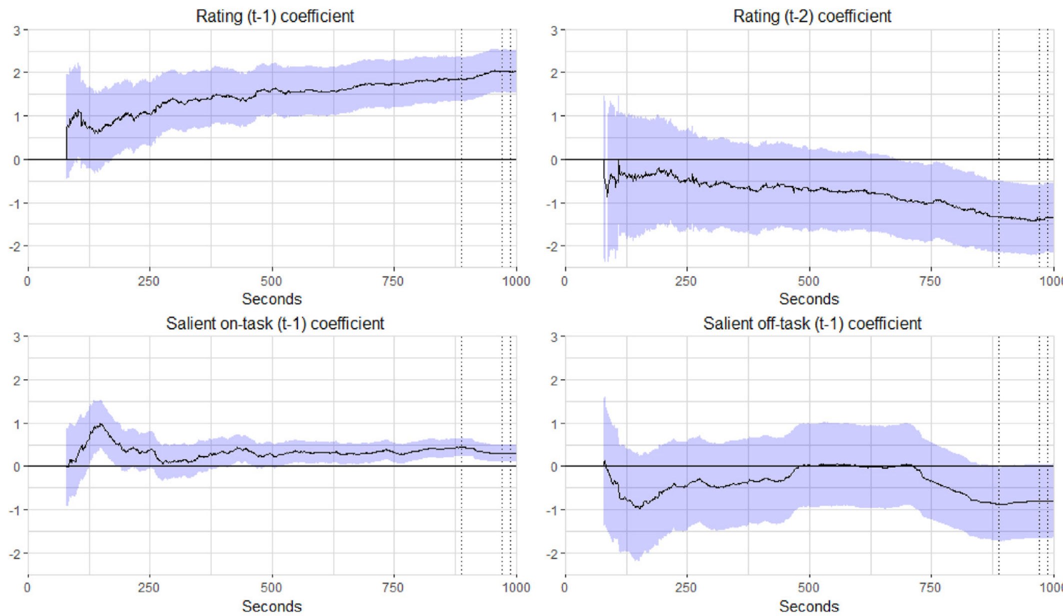

Fig. 4. Plots displaying structural breaks for the different coefficients, with blue areas indicating the 95% CIs and vertical dotted lines indicating time points when no further data from one video was added. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

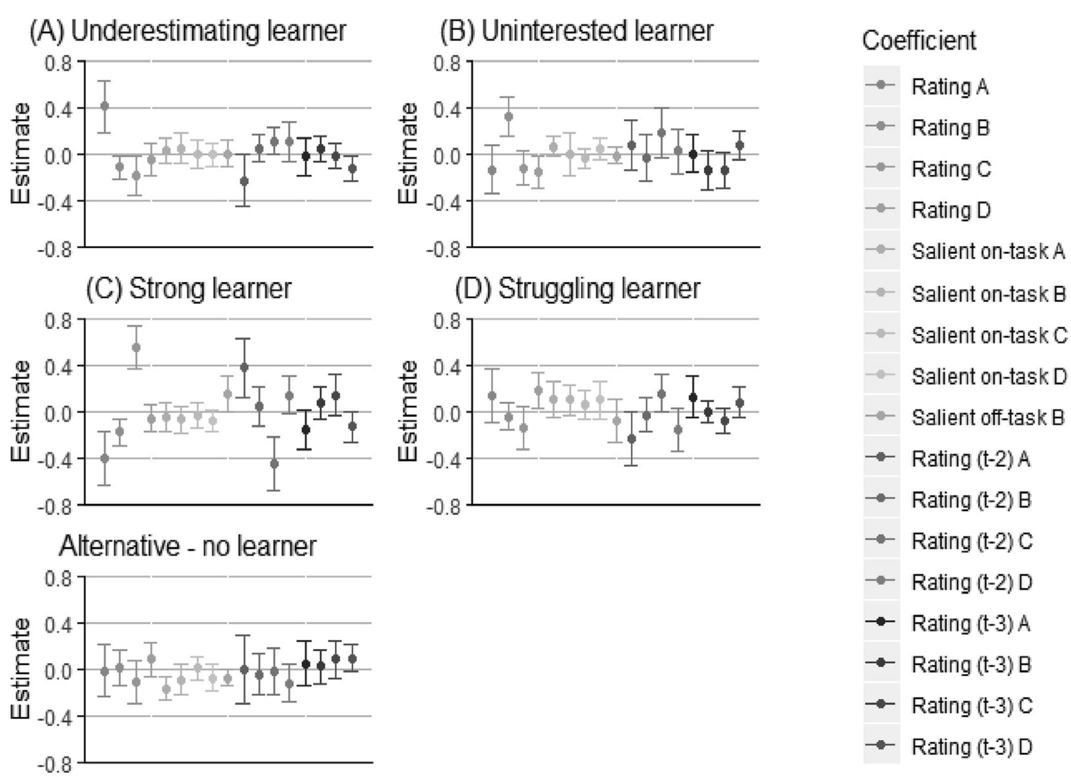

Fig. 5. Linear probability models for each learner and the alternative with 95% CIs. Outcome is equal to one if the respective learner was looked at or (for the alternative) no learner was looked at and zero otherwise. In the legend, A indicates the underestimating learner, B the uninterested learner, C the strong learner, and D the struggling learner.

sustained the course of instruction and tried to avoid misbehaving learners. These findings are particularly interesting, as research with stationary eye-trackers has demonstrated that novice teachers' attention is attracted by disruptive behavior and rather salient features (i.e., bottom-up influences) when watching a video rather than teaching themselves (Wolff et al., 2016). This mismatch is in line with research on people's gaze behavior finding different patterns in laboratory and real-world settings (Foulsham et al., 2011). According to Foulsham et al. (2011), these differences might be influenced by predictions of how the scene in the real world will change and the requirement to engage with the given task (i.e., top-down processes). The difference between our results and those of Wolff et al. (2016) could indicate, for example, that novice teachers' attention might be driven by the demands of the context (i.e., a more passive context when watching a video without the need to interact with learners vs. actual teaching in which they must interact with learners) and their underlying intention (i.e., observing a scenario vs. conveying learning content). Early research already found that novice teachers demonstrate certain inflexibilities when it comes to deviations from lesson plans (e.g., Livingston & Borko, 1989; Westerman, 1991). Therefore, top-down processes (i.e., plans and intentions) related to following their instructional agenda might guide novice teachers' attentional focus rather than top-down processes related to steady monitoring and the identification of problematic behavior. Nevertheless, it is important for inexperienced teachers to overcome the urge to focus mainly on actively engaged students and instead monitor the classroom evenly, as they have to identify inattentive students in order to encourage their active participation and support engagement and learning from all students (Seidel et al., 2020). By being more likely to react to salient behavior than rather unobtrusive cues, novice teachers might fail to identify students who need special attention because they are struggling (low-performing profile) and/or lack confidence in their skills (underestimating profile). Furthermore, it is important for teachers to be able to identify the underlying reasons for student behaviors, as a low-performing student needs different kinds of support than a student who underestimates him- or herself or a student who is simply not interested in the learning topic. Running the regressions separately for each preservice teacher revealed that the learners' salient behavior generally affected the preservice teachers' attentional focus. However, we found variations among individuals. For most preservice teachers in our sample, salient behavior exerted a positive effect on their attentional focus, meaning that they focused their attention on conspicuous rather than unobtrusive cues. However, one preservice teacher's attention was affected in the opposite way. Furthermore, while learners' behavior exerted a positive effect on preservice teachers' attentional focus, this effect was insignificant for one preservice teacher. This disparity is in line with previous findings (Stürmer et al., 2017; Dessus et al., 2016) implying different processes of attention allocation and indicating varying stages of schema construction in preservice teachers. For example, trying to avoid focusing on salient behavior could indicate top-down, rather than bottom-up, processes of attention allocation, as attention is not guided by striking cues, but by the intention to avoid this kind of behavior and focus on more subtly acting learners.

We assumed that the time course would influence how preservice teachers distribute their attentional focus during instruction. We found no effect of instruction time on the relationship between learners' behavior and preservice teachers' attention in general. Learners' behavior guided preservice teachers' attentional focus throughout the time course. At all times, teachers were more likely to focus on actively engaged learners compared with rather passive or even disturbing behavior. Furthermore, preservice teachers focused on actively engaged learners who exhibited salient behavior, especially during the second half of the instructional time. This might be due to preservice teachers' intention to convey certain learning content during the instructional period. When experiencing pressure to finish in time, they might pay more attention to learners who can help them pursue their goals, and thus focus their attention on students who display salient on-task behavior. Finally, the behavioral rating 2 s earlier exhibited a negative effect during the last third of the instructional time, indicating that preservice teachers' gaze is not stable. This is in line with previous research showing that novice teachers' attention while teaching is dominated by the short term, involving quick changes between AOIs (Stürmer et al., 2017). On the other hand, it might also indicate that the novices attempted to monitor the classroom after they had some time to get accustomed to the situation. With respect to salient off-task behavior, our results indicate no significant effect and rather unstable estimations. One explanation from a technical point of view might be the comparatively fewer data points considered salient off-task behaviour, which made the estimations less precise. A more content-based explanation would be that preservice teachers did not react as deliberately and consistently to salient off-task behaviour as they did to salient on-task behaviour. Thus, the estimations were rather imprecise because the preservice teachers reacted in a non-systematic way when salient off-task behaviour occurred.

Previous research indicates that students exhibit different kinds of observable behavior depending on their individual characteristics (Jurik et al., 2013; Pauli & Lipowsky, 2007) and that teachers generally prefer to interact with actively engaged students (Lipowsky et al., 2007). Therefore, we investigated whether profile-specific effects exist that guide preservice teachers' attentional focus. We found that learners' ratings exerted a generally positive effect. For example, asking questions increased the probability of being in the preservice teacher's attentional focus for all profiles. However, this effect was greatest for strong learners and weakest for struggling learners. This finding highlights a particular issue, as struggling students particularly need their teachers' attention. When teachers overlook students who are experiencing difficulties in understanding instruction, they fail to engage these students in the learning process, resulting in decreased and/or unsuccessful learning. Additionally, preservice teachers' attentional focus was affected differently by different profiles. For example, when the struggling learner participated more actively, only the uninterested learner's probability of being in the preservice teacher's AOI decreased, not those of the strong and underestimating learners. Moreover, only when the strong learner was participating actively did the probability of the preservice teacher continuing to look at him or her decrease. This might indicate that the preservice teachers knew that the strong learner was adequately engaged and were attempting to distribute their attention to other learners, as we did not find this effect with the other profiles. Taken together, our findings indicate profile-specific effects of learners' behavior on preservice teachers' attentional focus.

It should be noted that the number of data points in the category of salient off-task behavior was rather low, and the values less extreme compared with those in the salient on-task category (see Fig. 1). Whereas learners displayed actions from the upper extreme of the behavioral spectrum, like explaining content to fellow students, they did not engage in activities on the lower extreme of the scale, such as walking around and actively disturbing others or instruction. Such behaviour also occurs rather rarely in real classroom situations involving university students (Goldberg et al., 2019). Even though one of the learners was instructed to behave in an uninterested manner, the instructions for this learner included behaviors such as playing on their smartphone or sometimes disturbing their neighbor but not the whole group (i.e., passive and active off-task behavior but not interactive off-task behavior). However, our rating instrument has to cover the entire possible spectrum of learners' behavior in order to be considered valid. Thus, the observed patterns might be driven by too little variation in the displayed behavior, meaning that interpretations regarding preservice teachers' attentional focus with respect to salient off-task behavior should be drawn carefully.

Nevertheless, by using standardized situations, we were able to ensure comparable conditions for all participants involving a similar set of observable behaviors. Differences in the profile-specific behaviors were due to the preservice teachers' individual methods of interacting with the learners, as the learners were instructed to adapt to the situation naturally. Interestingly, even though all learners theoretically should have behaved in the same way, the variation in preservice teachers' interaction styles resulted in unequally pronounced behaviors.

Even though our sample size is rather small, the synchronization of the continuous data was based on almost 4000 s of material (i.e., data points). This constitutes a vast amount of data and — to the best of our knowledge — analysis of triangulated data like ours has never before been performed. Therefore, our study provides a promising starting point for systematically investigating interactions between teachers and students by deploying mixed methods and time series analysis. Our next step will be to explore the effects of real classrooms containing more students — and thus more demanding interaction processes — on novice teachers' attention. Furthermore, in future studies, it would be of great interest to compare experts and novices in order to identify knowledge structures and competencies that inexperienced teachers do not yet possess. Insights like these could have critical implications to help teacher educators and mentors train novice teachers.

6. Conclusion

Conducting a time series analysis of teachers' eye-tracking data in combination with continuous ratings of student behavior is a promising approach to analyzing teacher-student interactions during instruction in more detail. We found that inexperienced teachers are more likely to focus their attention on students who exhibit actively engaged behavior compared with rather passive or even disruptive behavior, and that this effect is stable across the period of instruction. Our findings further support the distinction between on-action and in-action research, as novice teachers in particular might behave differently when faced with the demands of actual classroom instruction and interaction.

However, the rating procedure for such synchronized data is time-consuming. To further study such interaction processes with larger sample sizes and in real classroom settings in which teachers usually teach more than four learners, automated assessment seems to be a promising next step (Goldberg et al., 2019).

Author statement

Patricia Goldberg: Conceptualization, Writing - Original Draft, Project administration, Formal analysis; Jakob Schwerter: Methodology, Formal analysis; Tina Seidel: Resources, Conceptualization; Katharina Müller: Resources, Conceptualization; Kathleen Stürmer: Supervision, Writing - Review & Editing, Conceptualization.

Author note

Patricia Goldberg and Jakob Schwerter are doctoral students at the LEAD Graduate School & Research Network [GSC1028], which was funded within the framework of the Excellence Initiative of the German federal and state governments. This work was supported by the German Research Foundation (Deutsche Forschungsgemeinschaft, DFG), grant number PR 473/6-3.

Appendix A

To see whether our results were driven by just one preservice teacher and cannot be generalized, we ran the multinomial regressions separately for each preservice teacher. However, not enough variation in salient off-task behavior existed to be estimated in the multinomial regression for Preservice Teachers 1 and 4. Thus, in our regression results, we did not differentiate between salient on- and off-task behaviors; instead, we used only one dummy variable indicating salient behavior in general (Figure A1). Regression results are displayed in Table A2. The rating is positively significant for Preservice Teachers 1, 2, and 3. Additionally, salient behavior is positively significant for Preservice Teachers 1, 3, and 4; however, it is negatively significant for Preservice Teacher 2. The second time lag of the rating is negatively significant only for Preservice Teachers 1 and 3.

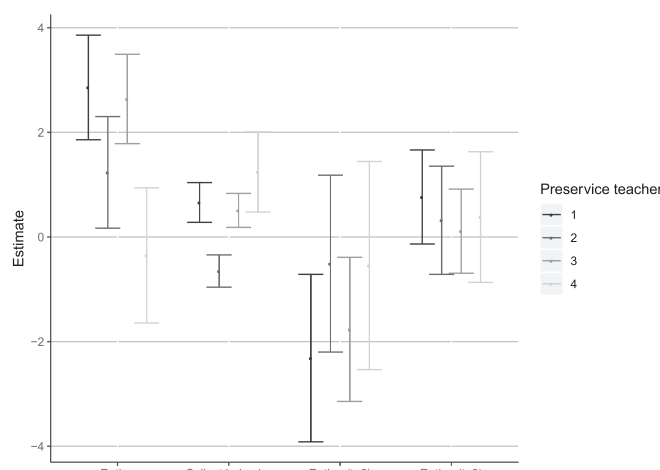

Fig. A1. Teacher-specific regressions for all preservice teachers with one variable indicating salient behavior in general, as the model could not be estimated for Preservice Teachers 1 and 4 otherwise. Whiskers indicate 95% *CI*s.

Table A2
Teacher-specific regression for all preservice teachers with one variable indicating salient behavior in general, controlling for teacher events, teachers, and time trend.

	Preservice teacher 1			Preservice teacher 2			Preservice teacher 3			Preservice teacher 4		
	b	SE	p	b	SE	p	b	SE	p	b	SE	p
Rating $_{t-1}$	2.86	0.51	<.001	1.23	0.54	.023	2.64	0.44	<.001	−0.35	0.66	.593
Salient behavior $_{t-1}$	0.66	0.19	.001	−0.65	0.16	<.001	0.51	0.17	.002	1.24	0.39	.001
Rating $_{t-2}$	−2.32	0.82	.005	−0.51	0.86	.554	−1.77	0.70	.012	−0.55	1.02	.591
Rating $_{t-3}$	0.77	0.46	.096	0.32	0.53	.544	0.11	0.41	.786	0.38	0.64	.550
x^2		796*			305*			373*			64*	
Pseudo R^2		0.2565			0.1095			0.1277			0.031	
Observations		1011			918			929			760	

Note: x^2 refers to the likelihood ratio test. We calculated McFadden's Pseudo R^2. $*p < .001$.

To investigate the specification including salient off-task behavior, we calculated multinomial regressions for Teachers 2 and 3. The estimation results are summarized in Figure A3, which presents the estimated coefficients of the variables of interest and the respective 95% *CIs* (for exact values, see Table A4).

The rating is significant for both teachers. Salient on-task behavior is negatively significant for Preservice Teacher 2, but positively significant for Preservice Teacher 3. Furthermore, salient off-task behavior is negatively significant for Preservice Teacher 2, but not significant for Preservice Teacher 3. The second lag of the rating is negatively significant only for Preservice Teacher 3. The third lag is statistically insignificant for both preservice teachers.

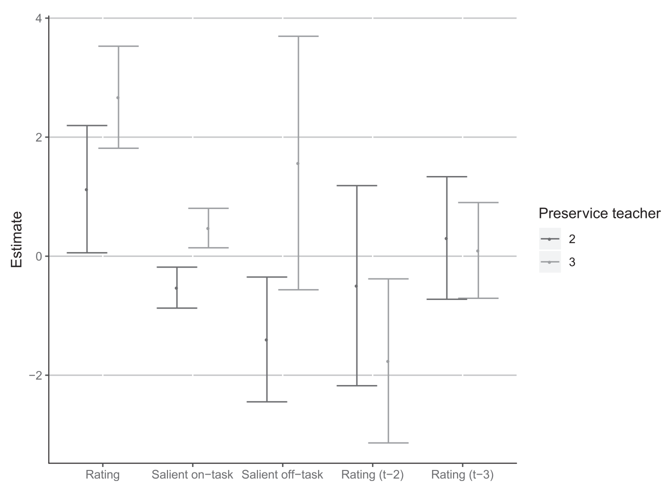

Fig. A3. Teacher-specific regression for Preservice Teachers 2 and 3, as only these have enough variation in the salient on- and off-task behavior variables. Whiskers indicate 95% CIs.

Table A4
Teacher-specific regression for Preservice Teachers 2 and 3, controlling for teacher events, teachers, and time trend.

	Preservice teacher 2			Preservice teacher 3		
	b	SE	P	B	SE	p
Rating $_{t-1}$	1.13	0.55	.039	2.67	0.44	<.001
Salient on-task behavior $_{t-1}$	−0.53	0.18	.003	0.47	0.17	.005
Salient off-task behavior $_{t-1}$	−1.40	0.54	.009	1.57	1.09	.150
Rating $_{t-2}$	−0.50	0.86	.564	−1.76	0.70	.012
Rating $_{t-3}$	0.31	0.53	.561	0.10	0.41	.813
x^2		307*			374*	
Pseudo R^2		0.1105			0.128	
Observations		918			929	

Note: x^2 refers to the likelihood ratio test. We calculated McFadden's Pseudo R^2. $*p < .001$.

Appendix B

Linear probability models for each learner and the alternative with outcome equal to one if the respective learner was examined or (in the alternative model) no learner was examined and zero otherwise. Controlled for teacher events, teachers, and time trend.

	(A) Underestimating learner			(B) Uninterested learner			(C) Strong learner			(D) Struggling learner			Alternative – no learner		
	b	SE	p	b	SE	p	b	SE	p	b	SE	p	b	SE	p
Rating A $_{t-1}$	0.42	0.12	<.001	−0.14	0.11	.207	−0.40	0.12	.001	0.14	0.12	.254	−0.01	0.12	.926
Rating B $_{t-1}$	−0.12	0.05	.024	0.32	0.09	<.001	−0.18	0.06	.002	−0.04	0.06	.465	0.01	0.08	.859
Rating C $_{t-1}$	−0.19	0.09	.027	−0.12	0.07	.115	0.56	0.10	<.001	−0.14	0.10	.136	−0.11	0.09	.232
Rating D $_{t-1}$	−0.05	0.07	.488	−0.16	0.07	.023	−0.06	0.06	.362	0.18	0.08	.023	0.09	0.07	.235
Salient on-task behavior A $_{t-1}$	0.03	0.06	.566	0.07	0.05	.124	−0.05	0.07	.492	0.11	0.08	.164	−0.17	0.05	.001
Salient on-task behavior B $_{t-1}$	0.05	0.07	.447	0.01	0.09	.981	−0.07	0.06	.268	0.10	0.07	.119	−0.09	0.07	.175
Salient on-task behavior C $_{t-1}$	−0.01	0.06	.96	−0.04	0.04	.380	−0.03	0.06	.611	0.06	0.06	.320	0.01	0.05	.862
Salient on-task behavior D $_{t-1}$	−0.01	0.05	.934	0.05	0.05	.319	−0.07	0.05	.125	0.11	0.09	.214	−0.08	0.06	.198
Salient off-task behavior B $_{t-1}$	0.01	0.06	.907	−0.01	0.03	.741	0.15	0.08	.062	−0.08	0.10	.426	−0.07	0.03	.035
Rating A $_{t-2}$	−0.23	0.12	.049	0.08	0.11	.486	0.38	0.13	.004	−0.23	0.12	.046	−0.01	0.15	.997
Rating B $_{t-2}$	0.05	0.06	.398	−0.03	0.11	.760	0.05	0.09	.561	−0.03	0.08	.715	−0.04	0.09	.659
Rating C $_{t-2}$	0.12	0.06	.047	0.19	0.11	.085	−0.45	0.12	<.001	0.16	0.08	.058	−0.01	0.10	.922
Rating D $_{t-2}$	0.10	0.09	.244	0.03	0.10	.765	0.14	0.08	.08	−0.16	0.09	.091	−0.12	0.08	.143
Rating A $_{t-3}$	−0.02	0.08	.807	0.01	0.08	.989	−0.16	0.09	.066	0.13	0.09	.166	0.05	0.10	.603
Rating B $_{t-3}$	0.04	0.06	.450	−0.14	0.09	.095	0.08	0.07	.290	<.01	0.05	.994	0.02	0.08	.752
Rating C $_{t-3}$	−0.02	0.05	.722	−0.14	0.08	.092	0.15	0.09	.101	−0.08	0.05	.155	0.09	0.09	.317
Rating D $_{t-3}$	−0.13	0.06	.025	0.08	0.06	.221	−0.13	0.07	.044	0.08	0.07	.226	0.10	0.06	.103
F		17.289*			20.6691*			38.7244*			23.5791*			21.759*	
R^2		0.115			0.134			0.225			0.150			0.140	
Observations		3633			3633			3633			3633			3633	

Note: *$p < .001$. A indicates the underestimating learner, B the uninterested learner, C the strong learner, and D the struggling learner.

References

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

Berliner, D. C., Stein, P., Sabers, D. S., Clarridge, P. B., Cushing, K. S., & Pinnegar, S. (1988). Implications of research on pedagogical expertise and experience in mathematics teaching. In D. A. Grouws, & T. J. Cooney (Eds.), *Perspectives on research on effective mathematics teaching* (pp. 67–95). Reston, VA: National Council of Teachers of Mathematics.

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

Blomberg, G., Stürmer, K., & Seidel, T. (2011). How pre-service teachers observe teaching on video: Effects of viewers' teaching subjects and the subject of the video. *Teaching and Teacher Education, 27*(7), 1131–1140. https://doi.org/10.1016/j.tate.2011.04.008

Böheim, R., Knogler, M., Kosel, C., & Seidel, T. (2020). Exploring student hand-raising across two school subjects using mixed methods: An investigation of an everyday classroom behavior from a motivational perspective. *Learning and Instruction, 65*, Article 101250. https://doi.org/10.1016/j.learninstruc.2019.101250

Cameron, A., & Trivedi, P. (2005). *Microeconometrics: Methods and applications*. Cambridge, England: Cambridge University Press.

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. https://doi.org/10.1177/002248718803900306

Charness, N., Reingold, E. M., Pomplun, M., & Stampe, D. M. (2001). The perceptual aspect of skilled performance in chess: Evidence from eye movements. *Memory & Cognition, 29*, 1146–1152. https://doi.org/10.3758/BF03206384

Chi, M. T. H., Glaser, R., & Rees, E. (1982). Expertise in problem solving. In R. S. Sternberg (Ed.), *Advances in the psychology of human intelligence* (pp. 1–75). Hillsdale , NJ: Erlbaum.

Chi, M. T. H., & Wylie, R. (2014). The ICAP framework: Linking cognitive engagement to active learning outcomes. *Educational Psychologist, 49*(4), 219–243. https://doi.org/10.1080/00461520.2014.965823

Cortina, K. S., Miller, K. F., McKenzie, R., & Epstein, E. (2015). Where low and high inference data converge: Validation of CLASS assessment of mathematics instruction using mobile eye tracking with experts and novice teachers. *International Journal of Science and Mathematics Education, 13*(2), 389–403. https://doi.org/10.1007/s10763-014-9610-5

Dessus, P., Cosnefroy, O., & Luengo, V. (2016). "Keep your eyes on 'em all!": A mobile eye-tracking analysis of teachers' sensitivity to students. In K. Verbert, M. Sharples, & T. Klobučar (Eds.), *Vol. 9891. Adaptive and adaptable learning. EC-TEL 2016. Lecture notes in computer science* (pp. 72–84). Cham, Switzerland: Springer International Publishing.

Döring, N., & Bortz, J. (2016). *Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften*. Wiesbaden, Germany: Springer.

Doyle, W. (1986). Classroom organization and management. In M. C. Wittrock (Ed.), *Handbook of research on teaching* (3 ed., pp. 392–431). New York, NY: Macmillan.

Duffy, G. G., Miller, S., Parsons, S., & Meloth, M. (2009). Teachers as metacognitive professionals. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Handbook of metacognition in education* (pp. 240–256). New York, NY: Routledge.

Emmer, E. T., & Stough, L. M. (2001). Classroom management: A critical part of educational psychology, with implications for teacher education. *Educational Psychologist, 36*(2), 103–112. https://doi.org/10.1207/S15326985EP3602_5

van Es, E., & Sherin, M. G. (2002). Learning to notice: Scaffolding new teachers' interpretations of classroom interactions. *Journal of Technology and Teacher Education, 10*(4), 571–596. Retrieved from https://www.learntechlib.org/primary/p/9171/.

van Es, E., & Sherin, M. G. (2008). Mathematics teachers' "learning to notice" in the context of a video club. *Teaching and Teacher Education, 24*(2), 244–276. https://doi.org/10.1016/j.tate.2006.11.005

Foulsham, T., Walker, E., & Kingstone, A. (2011). The where, what and when of gaze allocation in the lab and the natural environment. *Vision Research, 51*(17), 1920–1931. https://doi.org/10.1016/j.visres.2011.07.002

Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School engagement: Potential of the concept, state of the evidence. *Review of Educational Research, 74*(1), 59–109. https://doi.org/10.3102/00346543074001059

Girard, J. M. (2014). CARMA: Software for continuous affect rating and media annotation. *Journal of Open Research Software, 2*(1), e5. https://doi.org/10.5334/jors.ar

Goldberg, P., Sümer, Ö., Stürmer, K., Wagner, W., Göllner, R., Gerjets, P., & Trautwein, U. (2019). Attentive or not? Toward a machine learning approach to assessing students' visible engagement in classroom instruction. *Educational Psychology Review*. https://doi.org/10.1007/s10648-019-09514-z

Goodwin, C. (1994). Professional vision. *American Anthropologist, 96*, 606–633. https://doi.org/10.1525/aa.1994.96.3.02a00100

Grossman, P., Compton, C., Igra, D., Ronfeldt, M., Shahan, E., & Williamson, P. W. (2009). *Teaching practice: A cross-professional perspective*. Retrieved December, 2nd 2019 from https://tedd.org/wp-content/uploads/2014/03/Grossman-et-al-Teaching-Practice-A-Cross-Professional-Perspective-copy.pdf.

Haataja, E., Moreno-Esteva, E. G., Salonen, V., Laine, A., Toivanen, M., & Hannula, M. S. (2019). Teacher's visual attention when scaffolding collaborative mathematical problem solving. *Teaching and Teacher Education, 86*, 102877. https://doi.org/10.1016/j.tate.2019.102877

Helmke, A., & Renkl, A. (1992). Das Münchener Aufmerksamkeitsinventar (MAI): Ein Instrument zur systematischen Verhaltensbeobachtung der Schüleraufmerksamkeit im Unterricht. *Diagnostica, 38*, 130–141.

Hofer, M. (1997). Lehrer-Schüler-interaktion. In F. E. Weinert (Ed.), *Enzyklopädie der Psychologie. Bd.3. Psychologie des Unterrichts und der Schule* (pp. 213–252). Göttingen, Germany: Hogrefe.

Hommel, M. (2012). Aufmerksamkeitstief in Reflexionsphasen - eine Videoanalyse von Planspielunterricht. *Wirtschaft und Erziehung, 1–2*, 12–18.

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

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

Karweit, N., & Slavin, R. E. (1981). Measurement and modeling choices in studies of time and learning. *American Educational Research Journal, 18*(2), 157–171. https://doi.org/10.3102/00028312018002157

Kelly, S. (2008). Race, social class, and student engagement in middle school English classrooms. *Social Science Research, 37*(2), 434–448. https://doi.org/10.1016/

j.ssresearch.2007.08.003

Kersting, N. (2008). Using video clips of mathematics classroom instruction as item prompts to measure teachers' knowledge of teaching mathematics. *Educational and Psychological Measurement, 68*(5), 845–861. https://doi.org/10.1177/0013164407313369

Kersting, N., Givvin, K. B., Thompson, B. J., Santagata, R., & Stigler, J. W. (2012). Measuring useable knowledge: Teachers' analyses of mathematics classroom videos predict teaching quality and student learning. *American Educational Research Journal, 49*(3), 568–589. https://doi.org/10.3102/0002831212437853

Kounin, J. S. (1970). *Discipline and group management in classrooms*. Oxford, England: Holt, Rinhart & Winston.

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

Lahaderne, H. M. (1968). Attitudinal and intellectual correlates of attention: A study of four sixth-grade classrooms. *Journal of Educational Psychology, 59*(5), 320–324. https://doi.org/10.1037/h0026223

Lei, H., Cui, Y., & Zhou, W. (2018). Relationships between student engagement and academic achievement: A meta-analysis. *Social Behavior and Personality, 46*(3), 517–528. https://doi.org/10.2224/sbp.7054

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. Retrieved from https://www.pedocs.de/frontdoor.php?source_opus=5489.

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. https://doi.org/10.1177/002248718904000407

Lizdek, I., Sadler, P., Woody, E., Ethier, N., & Malet, G. (2012). Capturing the stream of behavior: A computer-joystick method for coding interpersonal behavior continuously over time. *Social Science Computer Review, 30*(4), 513–521. https://doi.org/10.1177/0894439312436487

McIntyre, N. A., Mainhard, T., & Klassen, R. M. (2017). Are you looking to teach? Cultural, temporal and dynamic insights into expert teacher gaze. *Learning and Instruction, 49*, 41–53. https://doi.org/10.1016/j.learninstruc.2016.12.005

McKinney, J. D., Mason, J., Perkerson, K., & Clifford, M. (1975). Relationship between classroom behavior and academic achievement. *Journal of Educational Psychology, 67*(2), 198. https://doi.org/10.1037/h0077012

Moreno, F., Reina, R., Luis, V., & Sabido, R. (2002). Visual search strategies in experienced and inexperienced gymnastic coaches. *Perceptual & Motor Skills, 95*(3), 901–902. https://doi.org/10.2466/pms.2002.95.3.901

Pauli, C., & Lipowsky, F. (2007). Mitmachen oder Zuhören? Mündliche Schülerinnen- und Schülerbeteiligung im Mathematikunterricht. *Unterrichtswissenschaft, 35*(2), 101–124. Retrieved from https://www.pedocs.de/frontdoor.php?source_opus=5488.

Posner, M. I. (1988). Structures and functions of selective attention. In T. Boll, & B. Bryant (Eds.), *Master lectures in clinical neuropsychology* (pp. 173–202). Washington, DC: American Pschology Association.

Prieto, L. P., Sharma, K., Wen, Y., & Dillenbourg, P. (2015). The burden of facilitating collaboration: Towards estimation of teacher orchestration load using eye-tracking measures. In O. Lindwall, P. Häkkinen, T. Koschman, P. Tchounikine, & S. Ludvigsen (Eds.), *Vol. 1. Exploring the material conditions of learning: The computer supported collaborative learning (CSCL) conference 2015*. Gothenburg, Sweden: International Society of the Learning Sciences.

Putnam, R. T. (1987). Structuring and adjusting content for students: A study of live and simulated tutoring of addition. *American Educational Research Journal, 24*(1), 13–48. https://doi.org/10.3102/00028312024001013

Sabers, D. S., Cushing, K. S., & Berliner, D. C. (1991). Difference among teachers in a task characterized by simultaneity, multidimensionality, and immediacy. *American Educational Research Journal, 28*(1), 63–88. https://doi.org/10.3102/00028312028001063

Sacher, W. (1995). *Meldungen und Aufrufe im Unterrichtsgespräch. Theoretische Grundlagen, Forschungsergebnisse, Trainingselemente und Diagnoseverfahren*. Augsburg, Germany: Wissner.

Santagata, R., Zannoni, C., & Stigler, J. W. (2007). The role of lesson analysis in pre-service teacher education: An empirical investigation of teacher learning from a virtual video-based field experience. *Journal of Mathematics Teacher Education, 10*(2), 123–140. https://doi.org/10.1007/s10857-007-9029-9

Schütz, A. C., Braun, D. I., & Gegenfurtner, K. R. (2011). Eye movements and perception: A selective review. *Journal of Vision, 11*(5), 1–30. https://doi.org/10.1167/11.5.9

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., Schnitzler, K., Kosel, C., Stürmer, K., & Holzberger, D. (2020). Student characteristics in the eyes of teachers: Differences between novice and expert teachers in judgment accuracy, observed behavioral cues, and gaze. *Educational Psychology Review*. https://doi.org/10.1007/s10648-020-09532-2

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

Seidel, T., Stürmer, K., Schäfer, S., & Jahn, G. (2015). How preservice teachers perform in teaching events regarding generic teaching and learning components. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie, 47*(2), 84–96. https://doi.org/10.1026/0049-8637/a000125

Sherin, M. G. (2007). The development of teachers' professional vision in video clubs. In R. Goldman, R. Pea, B. Barron, & S. J. Derry (Eds.), *Video research in the learning sciences* (pp. 383–395). Mahwah, NJ: Lawrence Erlbaum.

Sherin, M. G., & van Es, E. (2002). Using video to support teachers' ability to interpret classroom interactions. In *Paper presented at the society for information technology & teacher education international conference*.

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

Star, J. R., & Strickland, S. K. (2008). Learning to observe: Using video to improve preservice mathematics teachers' ability to notice. *Journal of Mathematics Teacher Education, 11*(2), 107–125. https://doi.org/10.1007/s10857-007-9063-7

Steffensky, M., Gold, B., Holodynski, M., & Möller, K. (2015). Professional vision of classroom management and learning support in science classrooms - does professional vision differ across general and content-specific classroom interactions? *International Journal of Science and Mathematics Education, 13*(2), 351–368. https://doi.org/10.1007/s10763-014-9607-0

Stipek, D. (2002). Good instruction is motivating. In A. Wigfield, & J. S. Eccles (Eds.), *Development of achievement motivation* (pp. 309–332). San Diego, CA: Academic Press.

Stürmer, K., & Seidel, T. (2015). Assessing professional vision in teacher candidates-approaches to validate the observer extended research tool. *Zeitschrift für Psychologie, 223*(1), 54–63. https://doi.org/10.1027/2151-2604/a000200

Stürmer, K., Seidel, T., & Holzberger, D. (2016). Intra-individual differences in developing professional vision: preservice teachers' changes in the course of an innovative teacher education program. *Instructional Science, 44*(3), 293–309. https://doi.org/10.1007/s11251-016-9373-1

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

Theeuwes, J., Atchley, P., & Kramer, A. F. (2000). On the time course of top-down and bottom-up control of visual attention. In S. Monsell, & J. Driver (Eds.), *Control of cognitive processes: Attention and performance XVIII* (pp. 105–124). London, England: The MIT Press.

Thiede, K. W., Brendefur, J. L., Osguthorpe, R. D., Carney, M. B., Bremner, A., Strother, S., … 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

Westerman, D. A. (1991). Expert and novice teacher decision making. *Journal of Teacher Education, 42*(4), 292–305. https://doi.org/10.1177/002248719104200407

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-z

Yamamoto, T., & Imai-Matsumura, K. (2015). Teachers' gazes and awarness of students' behavior: Using an eye tracker. *Comprehensive Psychology, 2*(6). https://doi.org/10.2466/01.IT.2.6