

1 Through the eyes of the teacher - Multimodal exploration of expertise differences in the
2 perception of classroom disruptions

3 Mandy Klatt¹, Dr. Gregor Kachel^{1, 2}, Dr. Christin Lotz¹, & Prof. Dr. Anne Deiglmayr¹

4 ¹ Leipzig University

5 ² Max Planck Institute for Evolutionary Anthropology

6 Author Note

7 We received funding from QualiFond of University Leipzig. We have no conflicts of
8 interest to disclose. This article is based on data used at conference presentations
9 (DACH-Nachwuchsakademie, 2022; EARLI SIG 11, 2022; EARLI SIG 27, 2022).

10 Correspondence concerning this article should be addressed to Mandy Klatt,
11 Dittrichring 5-7, 04109 Leipzig. E-mail: mandy.klatt@uni-leipzig.de

12 Through the eyes of the teacher - Multimodal exploration of expertise differences in the
13 perception of classroom disruptions

14 **Introduction**

15 Managing classroom disruptions is a crucial aspect of effective classroom management
16 (Evertson, Weinstein, et al., 2006; Kounin, 2006). Accordingly, teachers must be able to
17 quickly notice and appropriately react to significant events in the classroom. This ability is
18 referred to as classroom professional vision (**goodwin2015?** professional; Sherin, 2007).
19 The process of professional vision can be divided into two main aspects: focusing on
20 relevant situations for learning and teaching (“noticing”) and applying knowledge to draw
21 appropriate conclusions in these situations [“knowledge-based reasoning”; Seidel and
22 Stürmer (2014)]. Therefore, the early visual perception of classroom disruptions is a key
23 component to effectively maximize students’ learning time and minimize classroom
24 interruptions. According to Kounin (2006), these important classroom management
25 strategies are called “withitness” and “overlapping” and can be summarized under the
26 concept of monitoring (Gold & Holodynski, 2017). Learning to develop such classroom
27 management skills is a demanding and complex task for student teachers (Wolff et al.,
28 2017). Research on teacher expertise showed that expert and novice teachers differ in their
29 ability to perceive classroom events, “[...] whereas only a few studies have focused on the
30 basal process of noticing, i.e. the recognition of possible disturbing situations”
31 (**grub2020processp?**). Mobile eye-tracking data can fill this research gap by providing
32 new insights in how expertise differences in teacher’s professional vision manifest in
33 teacher-student interactions (Lachner, Jarodzka, & Nückles, 2016; Wolff, Jarodzka, Bogert,
34 & Boshuizen, 2016).

35 Theoretical background

36 Professional competence in teaching profession

37 Competence is a common term in everyday school life. In most cases, it refers to the
38 acquisition of competencies by students and how to promote them and make them
39 effective. In the same way, the term competence can be applied to teachers. Teachers have
40 to be competent in things such as teaching, educating, evaluating or innovating. Therefore,
41 competencies form an important and practical contribution in current educational research
42 and teacher education (Klieme, Hartig, & Rauch, 2008). They replace the rigid canon of
43 knowledge that defined professional qualification for generations. In recent years, the
44 concept of competence has experienced an upswing and is in the focus of educational
45 research in particular (*ibid.*). To put it simply, the concept of competence in empirical
46 research deals with the development of human resources and the productivity of education
47 (cf. Klieme et al., 2008, p. 4). In order to be able to make a meaningful conceptualization
48 definition of competence, it is necessary to take a closer look at different of research have to
49 be considered in more detail.

50 In science, there is some agreement on the two opposing perspectives of competence
51 versus empowerment. The former means a broad, concept, whereas empowerment refers to
52 the different components of competence (cf. of competence (cf. Blömeke, Gustafsson &
53 Shavelson, 2015, p.5). These can be of a cognitive, conative, affective, or motivational
54 nature (*ibid.*). To determining professional qualification, competence is an irreplaceable
55 tool.

56 Competence is learnable and can thus be improved through targeted practice
57 (cf. Weinert 2001; cited in Blömeke et al. 2015, p. 5). This is one reason, among others,
58 why this concept is so relevant for empirical school and development research. Another
59 crucial step in clarifying the concept of competence is to understand that the framework of
60 competence recognizes the importance of everyday life situations (Blömeke et al., 2015).

61 Therefore scholars argue, a definition must start from an authentic “analysis of
62 occupational or societal situations and must take into account the tasks and associated
63 associated cognitive, conceptual, affective, and motivational aspects” (Blömeke et al., 2015,
64 p. 5). Through this comprehensive approach, many inconsistencies in the definition of the
65 term can be resolved. The definition of competence is now no longer a question of whether
66 it is a number of cognitive skills or a combination of cognition, conation, affect, or
67 motivation. It combines “complex intellectual characteristics together with
68 affect-motivation, which underlies observable underlying performances” (Blömeke et
69 al. 2015, p. 6). This is best defined as a process or even continuum (see Figure 1), which
70 consists of different levels and should be understood as a multidimensional construct with
71 stronger and weaker Domains (*ibid.*).

72 The term competence finds its place in a variety of disciplines as an assessment
73 standard of qualifications. In this context, the term professional competence has become
74 established in research, which defines competence in a specific professional field. Thus,
75 based on the assumption that competence is more than mere knowledge, professional
76 competence offers a way to make teacher success in schools tangible (Kunter et al., 2013).

77 The term “professional competence” is the application of the concept to working life,
78 particularly in highly complex and demanding professions, in which mastery of situations is
79 especially dependent on the interplay of knowledge, skills, attitudes, and motivation.” [cf.
80 Epstein and Hundert (2002); cited in Kunter et al. 2013, p. 807]. Several studies could
81 prove that this idea can be successfully transferred to the teaching profession
82 [cf. Goodmann et al., 2008; Oser, Achtenhagen, and Renold (2006), Achtenhagen &
83 Renold, 2006; Tannenbaum & Rosenfeld, 1994; cited in Kunter et al., 2013, p. 807].
84 Accordingly, a quality standard can be applied to assess the quality of a teacher. This
85 quality is elementary in the school context because the teacher factor most influences
86 student learning (Marzano, Marzano, & Pickering, 2003). Thus, the effectiveness of
87 education and the learning growth of individual students can be enhanced by the teacher.

88 Research has shown that this component can even offset the influence of an ineffective
89 school setting (Marzano et al., 2003). This effect places teachers at the center of ensuring
90 effective teaching and learning and underscores the importance of professional competence
91 in this career field.

92 **Classroom Management**

93 Teachers and students are in the classroom in a place that is always alive,
94 characteristic of its heterogeneity and the interests of different groups. These the
95 circumstances of a classroom are also characterized as multidimensional, simultaneous,
96 immediate, unpredictable, public, and shared (Doyle, 2011, p. 99). All of these
97 circumstances place a number of demands on teachers, which primarily involve planning
98 and monitoring activities (*ibid.*). “Order in classrooms is not a consequence of reactions to
99 misbehavior but a condition established and sustained by the way a teacher organizes and
100 guides a complex system of classroom activities” (Doyle, 1989, p.12). Thus, “good”
101 teachers create the conditions for effective teaching through pre-planned actions and
102 activities that they are always in control of while teaching. This enables them to maximize
103 work time while keeping disruptions of any kind to a minimum to facilitate learning for all
104 participants (Doyle, 1989).

105 However, lessons, and specifically classroom disruptions, are unpredictable as well as
106 multifaceted and require a high level of expertise and competence on the part of the
107 teacher to handle. This trained handling, as described above, can be summarized under the
108 aspect of professional competence. It is expressed, among other things, in knowledge about
109 classroom management (Grub, Biermann, & Brünken, 2020) and can be understood as a
110 manifold toolbox of teachers to find their way in the challenging environment of the
111 classroom.

112 Classroom management significantly influences the ability to perceive and

113 subsequently respond to classroom events and thus organize for disruption-free instruction
114 (Wolff, Bogert, Jarodzka, & Boshuizen, 2015). Marzano et al. (2003) research created a
115 foundational work in school and developmental research by manifesting that instructional
116 strategy, curriculum design, and classroom management are significant roles in effective
117 teaching. In this regard, research agrees that effective leadership and effective learning are
118 closely related and classroom management can be considered a key component of teaching
119 (Wolff et al., 2015). Therefore, it is often referred to as the central variable of successful
120 teaching (Helmke & Helmke, 2015). Good classroom management is thus the prerequisite
121 for learning to take place (Wolff et al., 2015). While the term suggests that it is mainly
122 about rituals, rules, or measures that are meant to maintain order within the classroom, it
123 is about much more than these two components.

124 An initial explanatory approach to classroom management was provided by Doyle
125 (1990, p.355) with his definition of teaching, which he described as a “cognitive activity
126 based on knowledge of the probable course of events in the classroom and the ways in
127 which actions affect situations.” From this, it is clear that in order to teach successfully, a
128 teacher needs knowledge of how to process as well as represent the full range of classroom
129 events. In addition, it is necessary to be able to respond or interact proactively and
130 effectively to potential events at best.

131 Ultimately, classroom management is composed of a set of multifaceted skills that
132 include classroom structure and atmosphere, instructional decisions made by the teacher,
133 and pedagogical as well as practical knowledge (cf. Wolff et al., 2015, p. 71). These skills
134 essentially determine teacher decisions, interactions, and exchanges in the course of
135 instruction (*ibid.*). If these very skills are well-developed, they guarantee a smooth and
136 trouble-free course of instruction. This is crucial to maximize students' active learning time
137 (Helmke & Helmke, 2015). It, in turn, has been shown to correlate positively with
138 achievement growth and learning interest in a class (Helmke, 2008), placing effective
139 classroom management at the center of empirical research. In principle, these skills can be

¹⁴⁰ developed in the course of teaching education, as well as subsequent professional
¹⁴¹ experience, which is why they constitute an essential component of teacher expertise (Wolff
¹⁴² et al., 2015).

¹⁴³ **Interdependence of classroom management.** In research, various models exist
¹⁴⁴ for approaching the concept of classroom management. However, they are very similar in
¹⁴⁵ their core statements. Therefore, it is merely a matter of preference which concept is
¹⁴⁶ consulted to explain connections and terms. In the following, the network of effects of
¹⁴⁷ classroom management according to Helmke et al. (2014) is used as a reference model in
¹⁴⁸ order to be able to understand the embedding of the concept accordingly.

¹⁴⁹ As was made clear in the definition, classroom management is a complex construct
¹⁵⁰ that attempts to accommodate the various modes and challenging characteristics of a
¹⁵¹ classroom. Helmke et al. (2014) therefore depicts classroom management as a complex web
¹⁵² of effects (see Figure 2) that serves the overarching goal of maximizing active learning time.
¹⁵³ This has a particularly high value in school and development research and is considered a
¹⁵⁴ predictor of learning success, which could be confirmed in a meta-study ($d = 0.39$; Hattie,
¹⁵⁵ 2013, p.219). Due to this diversity of aspects to be investigated, classroom management is
¹⁵⁶ a fundamental research topic.

¹⁵⁷ The concepts, strategies and techniques used to implement successful classroom
¹⁵⁸ management depend first and foremost on various teacher characteristics and are
¹⁵⁹ “interrelated with the quality of teaching and personal relationships” (Helmke & Helmke,
¹⁶⁰ 2015, p.5). It should not be forgotten that this space never functions in isolation and
¹⁶¹ independently of other influences. It is embedded in diverse contexts, especially in the
¹⁶² school and class context (Helmke & Helmke, 2015). But also the cooperation with parents,
¹⁶³ a healthy evaluation culture, as well as the professional and didactic context have a
¹⁶⁴ significant impact on how effectively a class can be led (Helmke et al., 2014).

¹⁶⁵ Classroom management itself is made up of various components which, according to

¹⁶⁶ Helmke et al. (2014), can be divided into four categories: Rules and Procedures, presence,
¹⁶⁷ Time Management, and Establishing Desirable and Reducing Undesirable Student
¹⁶⁸ Behavior. Rules are considered the non-plus ultra of proactive management in the
¹⁶⁹ classroom, to whose agreements students and teachers alike must adhere (Helmke &
¹⁷⁰ Helmke, 2015).

¹⁷¹ The importance of rules in the classroom has been known and researched for some
¹⁷² time. Doyle (1989) relied as early as the late 1980s on studies showing that effective
¹⁷³ classroom management is due to timely implemented rules [cf. Emmer, Evertson &
¹⁷⁴ Anderson, 1980; Emmer, Sanford, Clements, and Martin (1982); cited in Doyle, 1989,
¹⁷⁵ p. 15]. Although he also stresses that establishing a system of rules is a difficult task.
¹⁷⁶ Classroom rules are situation-dependent, as certain agreements only apply to certain phases
¹⁷⁷ of the lesson (for example, speaking quietly during group work; Doyle, 1989). Furthermore,
¹⁷⁸ order is always established jointly, in cooperation with the learners (Doyle, 1989).

¹⁷⁹ As a result, students significantly influence these rules by their willingness to follow
¹⁸⁰ them. Procedures are another tool to maintain the flow of the lesson, as recurring routines
¹⁸¹ provide learners with security and clarity (Helmke & Helmke, 2015). This inevitably
¹⁸² reduces the teacher's workload, eliminates additional explanations are no longer necessary,
¹⁸³ and it is clear to the learners at all times what needs to be done. In
¹⁸⁴ addition, teachers can present instructions for corresponding routines through non-verbal
¹⁸⁵ actions to save time (Helmke et al., 2014). For the reasons just explained, it is
¹⁸⁶ clear why Hattie (2013, p. 122) gives this component of classroom classroom management
¹⁸⁷ an effect size of $d = 0.76$.

¹⁸⁸ No other aspect of classroom management is as effective for classroom interruptions
¹⁸⁹ as the teacher's omnipresence (Hattie, 2013, p. 122); the effect size is $d = 1.42$. (Helmke &
¹⁹⁰ Helmke, 2015, p. 6). The behavioral pattern of presence is not a new concept. It was
¹⁹¹ already recorded by kounin1970discipline under the term *withitness*. Later research

192 confirmed and validated his assumptions and definition of the term (cf. Helmke et al.,
193 2014, p. 10). In order to be present, the teacher must be able to control the classroom
194 adequately and superficially effectively at all times. This means she must be able to
195 recognize relevant cues or situations that require possible action (Wolff et al., 2015).
196 Presence can consequently be illustrated very vividly by means of “eyes in the
197 back of the head” (Kounin, 2006). In addition to presence, Kounin (2006) uses the term
198 overlap. It is a kind of multitasking ability to be able to cope with parallel processes and
199 situations in the classroom. The two dimensions of overlap and presence can also be
200 summarized together under the term monitoring (Gold & Holodynski, 2017). When talking
201 about ongoing monitoring in the classroom, this process includes teachers’ awareness of
202 concurrent events as well as their demonstration of this to learners (Grub et al., 2020).
203 Well-developed monitoring behaviors enable teachers to detect disruptions early and
204 increase overall awareness of relevant events in the classroom (*ibid.*). From this point of
205 view, it becomes clear why awareness is a core component of teacher competence.

206 As a third component influencing the quality of classroom management, cites Helmke
207 et al. (2014) time use and management. This seems trivial at first, considering that
208 classroom management serves the goal of maximizing learning time anyway. However,
209 Helmke & Helmke (2015, p. 6) emphasize how elementary it is to identify and minimize
210 possible “time thieves”. This includes factors such as unpunctuality or slow transitions, as
211 well as the underemployment of learners (*ibid.*). If these take up too much space in the
212 classroom, less time is available for teaching important lesson content. This can be
213 prevented by teachers minimizing such “time thieves” in advance through well-planned
214 classroom management.

215 In order to enforce desired student behavior or to reduce contradictory behavior,
216 teachers can use various discipline-related interventions and also make use of negative
217 sanctions and positive reinforcement (Helmke & Helmke, 2015). The authors emphasize
218 that a healthy balance must be found when using these strategies. Over the years, this

219 constructive approach has fallen into disrepute due to its behaviorist character. However,
220 elements such as reinforcement, extinction or punishment are not negligible laws of learning
221 psychology whose effectiveness has been proven by Hattie (2013), among others. These
222 behavioral psychological measures have an effect size of $d = 0.76$ (cf. Hattie, 2013, p.215)
223 and are therefore an effective means of reducing disruptions (Helmke & Helmke, 2015).

224 Researchers in pedagogy also point out, however, that it is not only pure knowledge
225 about classroom management that determines its effectiveness and influences it (Helmke &
226 Helmke, 2015). The teacher personality, which includes characteristics such as subjective
227 tolerance margins, i.e. at what point does one perceive a certain behavior as disruptive,
228 also plays a role. This includes further aspects such as emotion control, empathy or the
229 credibility of a teaching person that arises through appearance. These characteristics,
230 which occur in many ways, have a direct impact on the quality of teaching and classroom
231 management, as they influence the planning behavior, perception and reaction of teachers.

232 **Professional vision**

233 Competence assessment plays a central role in the teaching profession and its
234 training. It follows the approach of concentrating on action-relevant as well as
235 lesson-related content in order to train corresponding competences for teaching practice
236 (KMK, 2004). In order for teachers in heterogeneous classrooms to be able to filter relevant
237 teaching aspects, they need knowledge about the control of their attention processes (V. L.
238 Barth, 2017). Science summarizes this competence under the so-called professional vision,
239 which is now a widely empirically studied research topic (Jahn, Stürmer, Seidel, & Prenzel,
240 2014). The concept of professional vision can originally be traced back to the
241 anthropologist Charles Goodwin and became known under the term Professional Vision
242 (Sherin, 2001). It provided the answer to Goodwin's question regarding the development of
243 a professional vision, in the context of the profession, compared to the vision of everyday
244 situations. Goodwin uses the comparison of an archaeologist and a layperson to

245 figuratively describe the differences. In doing so, he investigated which ability it is that
246 makes the archaeologist see and interpret a landscape differently than the layperson
247 (cf. Sherin, 2001, p. 75). What looks to the former merely like sand, rocks and stones, is to
248 the trained eye of the archaeologist a base of a column or architectural fragments.
249 Goodwin (1994, p.606) defines professional vision as “socially organized ways of seeing and
250 understanding events that are answerable to the distinctive interests of a particular social
251 group” (cited in Sherin, 2001, p. 75). It is assumed that one can see certain phenomena or
252 occurring events in a certain way due to one’s professional affiliation (Sherin, 2001).
253 Particularities are more easily recognized because the gaze becomes more trained over time
254 and thus profession-relevant features appear more distinctive. In the previous example of
255 the archaeologist, this would mean that he noticed variations in color, texture and sand
256 and saw a cluster of stones as a possible element of a larger structure (cf. Sherin, 2001,
257 p.75). This distinguishes experts significantly in their visions from non-experts in the field,
258 as they had learned to filter their environment for crucial features.

259 **Teachers’ professional vision.** Based on the assumption that the professional
260 vision can develop over time, Sherin (2001) postulates that teachers can also acquire a
261 professional vision. As teachers develop from novices to experts in their field, they
262 accumulate expertise in a variety of areas (cf. Sherin, 2001, p. 76). This can be seen, among
263 other things, in flexible reactions to unexpected events or the smooth implementation of
264 routines in the classroom Leinhardt & Greeno (1986). This type of action can be described
265 as a skill that helps teachers make decisions about what is relevant and what is not within
266 the complex classroom setting (van Es & Sherin, 2002). Such decisions are based on
267 professional knowledge that guarantees effective teaching and learning in actual classroom
268 situations (Jahn et al., 2014). Professional awareness is therefore referred to as an indicator
269 of how knowledge is represented in teachers (Stürmer et al., 2017). It can also be seen as a
270 bridge or link between knowledge and action, or professional competence and behavior
271 (**grub2020process0?**). It is important to understand that professional vision is acquired

and not inherent. Rather, it is subject to a process of acquisition that is constantly progressing as teachers are exposed to events and learn to interpret them (Wolff et al., 2016). Fundamentally, then, prospective teachers must select what to focus their attention on in the classroom and how to understand these events (Stürmer et al., 2017). Finally, teaching and learning take place in a complex environment. Simultaneous events and situations cannot be given the same attention. “Selective perception is necessary in order to distinguish important from unimportant features and to focus attention on relevant aspects” (V. L. Barth, 2017, p. 19). Furthermore, teachers benefit from a trained perception of problems, as this is a significant factor in being able to preventively counteract preventive measures (Wettstein, 2013).

Basically, the complex process of professional vision is divided into three sub-processes (Grub et al., 2020). First of all, the identification, the recognition, of an event takes place. Then connections are made, which are finally evaluated by the teacher using his or her professional knowledge. In research, these processes are often combined and one speaks of noticing to describe recognition (Grub et al., 2020). Knowledge-based reasoning includes the last two processes of applying knowledge to make appropriate decisions (*ibid.*). During this process, it becomes clear that the recognition of relevant situations, such as disruptions, precedes all others. This process is therefore considered necessary, without which professional vision cannot take place (Grub et al., 2020). As described in the chapter on competence, perceptions is also of particular importance in the competence model according to Blömeke et al. (2015). This is because the process of perception forms an essential basis for situation-specific skills of teachers (Grub et al., 2020). Inferentially, recognition is inevitable for the perception of classroom events and therefore highly relevant for effective classroom management (*ibid.*).

Even though the recognition of relevant situations in the classroom is the starting point for all further subsequent actions, the processes of knowledge-based reasoning within professional vision are not to be neglected. It is also said that this process is an indicator of

299 the quality of how knowledge is applied to a given situation (Grub et al., 2020), as teachers
300 need to make knowledge-based inferences. Qualitative research on knowledge-based
301 reasoning concluded that it can be subdivided into three further sub-processes. Sherin
302 (2007; cited in V. L. Barth, 2017, p. 20) made a first subdivision, which consists of the
303 processes of describing, evaluating and interpreting. When a teacher describes what she
304 notices, she concentrates on observable characteristics, i.e. important aspects of teaching
305 (*ibid.*). When evaluating, statements are made about the quality of the teaching actions,
306 and when interpreting, conclusions are drawn about possible mechanisms of the event (cf.
307 V. L. Barth, 2017, p. 19). Seidel et al. (2010), on the other hand, focus on
308 German-speaking countries in their research on professional vision and modified Sherin's
309 original concept. They also divide knowledge-based-reasoning into three levels, but these
310 differ: describing, explaining and predicting (V. L. Barth, 2017). More precisely, these
311 processes encompass the ability to differentially describe instructional components that
312 have an impact on learning; to make theoretically sound explanations about instructional
313 situations; and to predict the effects of these events in relation to the teaching-learning
314 process (cf. Seidel et al., 2010, p. 297). This scientific work examines only the processes of
315 noticing, which is why these two approaches have been presented for the sake of
316 completeness. However, it is not useful to specify which classification of knowledge-based
317 reasoning is referred to in arguments. is referred to in argumentation.

318 Against this scientific background, it can be deduced that the two processes of
319 noticing and knowledge-based-reasoning in professional vision cannot be considered
320 separately from each other, but that they have a reciprocal effect
321 (**sherin2007development7?**; cited in **barth2017professionelle7?**). What is considered
322 and perceived as relevant in a teacher's selective vision (noticing) functions as the basis of
323 the subsequent processes of describing, evaluating, and interpreting
324 (knowledge-based-reasoning). Conversely, by engaging with these processes, the teacher's
325 perception is sensitized. This leads to the fact that in the future certain characteristics of

326 this situation can be recognized rather.

327 **Measuring professional vision: eye-tracking technology in research**

328 Professional vision of teachers is a complex undertaking, consisting of various
329 cognitive and visual processes. In the previous chapter, the importance of perception and
330 its associated visuality was emphasized, which places the eye at the center of
331 investigations. In research, process-based methods are used to make these perceptual
332 processes visible. In this context, eye-tracking (ET), a modern method of recording gaze,
333 has proven to be an effective tool in research. It allows us to literally see through the eyes
334 of recipients and provides insight into another perspective. Intrinsic human cognitive
335 processes are often compared to a black box in science, as these activities are not directly
336 measurable (Rakoczi, 2012). Until now, cognitive processes could only be studied by
337 verbalizing the respective subjects. However, this method is highly susceptible to
338 interference due to its incomplete accessibility (Dessus, 2016). The increasingly popular
339 research method of eye-tracking technology therefore represents an innovative means of
340 externalizing, i.e., making visible, internalized processes by means of eye movements
341 (Gegenfurtner et al., 2018) Stimuli in the environment are always perceived by our eyes as
342 the first receiver. They take in the information and pass it on to our brain so that it can
343 finally be processed there. By measuring this complex process, eye movements can be
344 recorded and subsequently analyzed (Rakoczi, 2012). Information is obtained about the
345 visual behavior, attentional focus, selection decisions, and even learning processes of the
346 investigator (*ibid*). In addition to these qualitative results, which provide information
347 about underlying cognitive processes, quantitative data can also be collected through
348 eye-tracking (Cullipher, 2018). These provide information, for example, about how long
349 and in what order stimuli are looked at, or whether certain parts are looked at more often
350 than others (*ibid*.). This enables research to make data objectively assessable.

351 Eye-tracking technology can be roughly divided into two types of devices: Stationary

352 and mobile systems (Blake, 2013). In research so far, mainly eye-tracking devices have
353 been used, which phenotypically resemble a computer (*ibid.*). They are also referred to as
354 remote devices and, once calibrated, measure eye movements in a non-contact manner
355 using infrared light (**egenfurtner2018mobiles?**). However, this technique reaches its
356 limits when it comes to collecting data in everyday situations. For this purpose, the more
357 flexible variant, mobile eye tracking (MET) or head-mounted or wearable eye trackers, can
358 be used. Typically, the technology is found as a head-mounted device on helmets or
359 integrated into an eyeglass frame (Blake, 2013). The difference with stationary tracking is
360 that in addition to an eye camera and infrared light source, additional cameras are built in
361 to capture head movements (*ibid.*). These are intended to help document the subject's field
362 of view. Therefore, such systems are particularly suitable for applied sciences, as is the case
363 in empirical school and development research (Gegenfurtner et al., 2018).

364 Eye-tracking research is based on several fundamental assumptions. The connection
365 between mental processing and eye movements has already been carefully investigated and
366 provides viable empirical evidence to date (cf. Blake, 2013, p. 380). In this regard,
367 hoffman1995role obtained the finding that when the eyes are focused on an object, this is
368 associated with an attentional focus of the person. Based on these findings, ET technology
369 is based on the following two assumptions: The immediacy assumption and the eye-mind
370 assumption (cf. Rakoczi, 2012, p.87). The former defines that cognitive evaluation
371 processes begin immediately upon fixation on an object (Cullipher, 2018). If a new
372 stimulus attracts attention and is fixated, the observer starts processing this new
373 information (*ibid.*). The eye-mind hypothesis assumes that whatever is fixated by the eye,
374 is subsequently processed by the mind (*ibid.*). It thereby expansively defines, that objects
375 are visually focused only until they are cognitively evaluated or until there is have been
376 cognitively evaluated or there is still an interest (cf. Rakoczi, 2012, p.87). However, a
377 situational context must exist for the assumptions just explained to be are valid (*ibid.*).
378 Through these functions, it becomes clear that the intrinsic processes of recognition in

379 professional vision can be measured by eye-tracking technology.

380 Despite the externalization of internalized processes and the resulting multitude of
381 data potentially to be collected, ET technology cannot provide answers about the exact
382 thoughts during stimulus processing. ET can identify where and when people are looking,
383 but it does not answer why. For this, it is helpful to triangulate the data and apply
384 alternative elicitation (Cullipher, 2018). One proven method is to conduct reflective
385 interviews, which can be conducted in the format of a think aloud protocol (Konrad, 2020).
386 This thinking aloud opens up the possibility of gaining insights into subjects' thinking,
387 feeling, and intentions (edb.). An elaborated form of the Think Aloud Protocol is a
388 Stimulated-Recall Interview (SRI). This is a qualitative scientific method to obtain data on
389 preactional thought processes (cf. Messmer, 2015, p.4). The pre-recorded lesson sequences
390 are played back and subjects have the opportunity to express themselves about their
391 "thoughts, beliefs, norms, and decisions" (Messmer, 2015, p.4) during the interview
392 process. Thus, the visual data of the ET can be matched with cognitive processes and
393 processes of knowledge-based-reasoning can be better understood. Individual-dependent
394 thinking is now no longer pure speculation and can be made objectively observable in a
395 way (Messmer, 2015). Subjects also have the opportunity to explain their actions and
396 become aware of the basis for their decisions (ibid.).

397 **Parameters and indicators of teachers' professional vision.** In university
398 education, there is an increasing focus on developing the professional vision of future
399 teachers (Stürmer et al., 2017). In order to determine the professional vision of teachers by
400 means of mobile eye-tracking (MET), different indicators are worth considering. This
401 scientific work focuses primarily on processes of noticing, i.e., the identification of relevant
402 teaching situations as well as perception time. Therefore, in the following we will focus
403 more on the parameters and indicators necessary for these processes.

404 In MET, collected data sets are usually output according to different parameters:
405 Fixations and Saccades (Gegenfurtner et al., 2018). The latter describes gaze jumps, i.e.,

406 movements of both eyes, which are useful to capture new fixations (Eye-Tracking
407 Kompetenzzentrum Schweiz, 2021). Significant for this scientific work and the analysis of
408 professional vision are the fixations as well as subjects' areas of interests. Fixations
409 describe the period of time in which no movements are emitted by the eye (Holmquist et
410 al. 2011) and are based on the eye-mind hypothesis described in the previous chapter. A
411 fixation of an object can be characterized according to its duration, frequency, and
412 distribution. In the illustration, the radius of a circle defines the duration of a fixation
413 (Gegenfurtner et al., 2018). These data can also be used to describe perception times, such
414 as the time until the first fixation on a particular object. In order to make this time period
415 objectively measurable make it objectively measurable, a concept developed by Tobii Pro
416 Lab Analyzer Software is used. In this software these organized recorded data are called
417 Time of Interests (short TOIs), which describe time intervals of corresponding events or
418 behaviors (Tobii Pro, 2021). Specifically, they are defined by the beginning and end of the
419 corresponding event.

420 By means of eye movement measurement, attention allocations to areas of interest or
421 objects are made accessible in addition to pure fixations (Gegenfurtner et al., 2018). In
422 research, these regions in the field of view of subjects are referred to as areas of interest
423 (AOI) (edb.). They are significant in that, for the measurement of perceptual velocities of
424 a relevant instructional event, not only TOIs but also the locations of fixations are crucial.
425 Thus, it can be determined how subjects distribute their attentional resources and whether
426 they recognize and perceive relevant events.

427 **Expertise differences in professional vision.** Since it is assumed that
428 professional vision is a preferably knowledge-based process, differences between experienced
429 and inexperienced teachers have already been identified. These differences were in favor of
430 experts, since a connection between teaching experience and the associated increase in
431 knowledge acquisition is assumed (Grub et al., 2020). Student teachers do not have
432 well-structured experiential knowledge due to their almost non-existent practice in

everyday school life (Wolff et al., 2016). This causes them selection difficulties in perception and is called pedagogical knowledge gap (cf. Wolff et al., 2016, pp. 2016, p.260). Through newly gained experiences, the professional vision of a teacher is continuously restructured and updated, and mental representation is constantly realigned accordingly (ibid.). This has been proven in research as a positive correlation of input to teaching and learning theories in studies and professional classroom perception has been found (Stürmer et al., 2012). Therefore, it can be assumed that there is a relationship between teacher expertise, their visual processing, and mindfulness for classroom management (Wolff et al., 2016). Research has also shown that declarative knowledge acquisition, i.e., acquisition of factual knowledge, supports perceptual processes [Seidel et al., 2012; cited in Treisch (2018), p.26] Accordingly, this increased knowledge acquisition would explain a positive impact on professional teaching perception. Among the most important prerequisites for this acquisition, baumert2006key include prior knowledge and interest. The former is elementary for new knowledge acquisition about effective teaching and learning, as it supports processing, integration, and consolidation of it (Stürmer, 2011). Furthermore, it should be noted that teachers who have developed schemas through teaching experience also exhibit good professional teaching perceptions (cf. Sabers, Cushing, & Berliner, 1991; cited in Treisch, 2018). This is explained by increased filtering mechanisms that support perceptual processes, i.e., recognition. Specifically, it has been shown that experienced teachers have more fixations on relevant areas because they have already learned theories on how to detect potential clutter in the classroom (Grub et al., 2020). Furthermore, experts have an advantage when it comes to coding information. Novices, on average, have a longer viewing times, so-called fixation durations, which is due to their still weakly developed expert knowledge (ibid.).

Mobile eye tracking in teaching-learning situations. As explained in detail earlier, teachers face a number of challenges during teaching. The teacher as organizer has to permanently scan the classroom for potential interruptions or possible events occurring

460 such as classroom disruptions. In the worst case, they force a standstill in the flow of
461 instruction, which is naturally to be avoided. Lohmann (2015, p.13) describes classroom
462 disruptions as “events that impair, interrupt, or make impossible the teaching-learning
463 process by partially or completely overriding the conditions under which teaching and
464 learning can first take place.” However, such an event only becomes a disruption when
465 actors inside the classroom, perceive, interpret and finally evaluate it as disruptive [cf.
466 Eckstein, Grob, and Reusser (2016); cited in, Eckstein, 2018, p.11]. Thus, instructional
467 disruptions are strongly dependent on perception and involve elements of production as
468 well as reception (*ibid.*). These discipline conflicts can be divided into four categories of
469 disruptive student behavior: Verbal disruptive behavior (chattering, heckling), lack of
470 eagerness to learn (mental absence in the form of drawing or playing on the cell phone),
471 physical disruptions (snapping, drumming, pen clicking), and aggressive behavior (temper
472 tantrums, attacking people; Lohmann, 2015).

473 In order to be able to recognize inappropriate behavior in a timely manner, the
474 teacher must have a pronounced omnipresence, as well as pedagogical-psychological
475 knowledge about teaching and learning processes (Cortina et al., 2018). These skills and
476 competencies are reflected in how teaching situations relevant to learning are perceived and
477 interpreted by teachers (*ibid.*). In this context, a multitude of visual attention processes
478 occur during classroom activities. This simply means the attention or the conscious
479 non-attention to visual stimuli (Rüth et al., 2020). This situational awareness of a teacher
480 and the accompanying control of his or her visual attention forms the foundation of
481 situation-specific skills (cf. Rüth et al. 2020). These data and insights potentially to be
482 collected form an interesting field of research and useful tool for problem-based teaching
483 education (Cortina et al., 2018).

484 In the last decade, visual input through recorded teaching situations has been
485 preferred (*ibid.*). Subjects are thus shown an outside perspective from different points of
486 view. This can happen either as self or external video and allows micro-analysis or

487 interaction sequences through the characteristic of repeated playing of a video (*ibid.*). This
488 characteristic makes it possible to analyze corresponding scenes more intensively and in
489 more detail. This results in a potential for reflection on teaching situations in teacher
490 education that should not be underestimated (cf. Rüth et al., 2020). The problem with
491 these videotaped sequences, however, is the lack of access to the first-person perspective
492 (*ibid.*). As a result, students usually evaluate and comment on their own videos more
493 negatively than those of others afterwards (Cortina et al., 2018). They find it difficult to
494 analyze the sequences objectively and to leave out unimportant details that concern their
495 person. However, this is not purposeful and conducive to the intended learning process of
496 the participating teachers. This is due to the fact that self-drawn cognition is activated in
497 self-videos and that when negative events occur, subjects tend to defend their actions
498 (*ibid.*). Kleinknecht and Schneider (2013) analyzed this process in more detail and found
499 that this is particularly problematic for the reflective learning process, as it has a
500 counterproductive effect on analyzing possible alternative actions. They also found that in
501 third-party videos, reflection on potential alternative actions was twice as frequent as in
502 self-recorded videos (*ibid.*)! In addition, recorded video sequences differ significantly from
503 real classroom settings in which the teacher can independently determine her field of view.
504 The observers cannot decide for themselves what to pay attention to and for how long to
505 pay attention to a situation or event (Sherin, 2001). Based on this finding, it was
506 significant for researchers to use an alternative elicitation technique to factor out the
507 negative effects of self and other videos.

508 A successful method to reduce this hindering factor of self-drawn cognition and still
509 use self videos for teacher education is to use recordings that show the teacher's perspective
510 and furthermore make it known where the teacher was looking (cf. Cortina et al., 2018). In
511 MET's video material, teachers can never see their teaching through their own eyes. This
512 makes it possible to focus more specifically on essential aspects, such as classroom
513 disruptions and interactions with learners. Although it has not yet been clearly

514 demonstrated that significantly less self-referential cognition was elicited, self-videos
515 counteract over-focusing on the acting person (cf. Cortina et al., 2018). Therefore, MET
516 represents a target-rich method for capturing teachers' attentional processes in the
517 classroom. Previous studies with large data sets on teachers' visual expertise have
518 predominantly used stationary ET (Wolff et al., 2016). In MET studies, the samples turn
519 out to be much smaller; here, the largest study to be mentioned is with 12 experienced and
520 12 inexperienced teachers by Cortina et al. (2015).

521 **Perceptions of classroom disruptions and their parameters.** The attentional
522 processes of teachers in the classroom that can be captured by MET are complex, as
523 already explained, and are related to various parameters. Individual personality
524 characteristics play a central role when it comes to the perception of relevant classroom
525 events. In general, it should be noted that novices have greater problems in classroom
526 management as well as in dealing with difficult student behavior (V. L. Barth, 2017).
527 Often, this is due to their late or even lack of awareness of classroom disruptions. However,
528 their lack of expertise is not the sole reason for this phenomenon. Individual beliefs and
529 convictions of teachers have a direct influence on the perception of a classroom situation
530 and are thus independent of the experience of the teacher (Baumert & Kunter, 2006) Thus,
531 in addition to teaching quality, emotions and emotional behavior of teachers can also be
532 cited as factors. Individual person characteristics as well as contextual factors not only
533 have an effect on teaching perception, but likewise on the sense of disturbance of a
534 situation (Eckstein, 2019). These specific characteristics include aspects such as self-image,
535 motivation, level of aspiration, attitudes, coping goals, and value orientations (Bakker &
536 Demerouti, 2006; cited in Wickeren, 2019, p. 33). Contextual characteristics include
537 didactic staging or general level of disruption. Undesirable behavior is perceived more
538 strongly by teachers when the class usually displays disciplined and orderly behavior
539 (Eckstein, 2018). This reference group effect is also known as the Big-Fish-Little-Pond
540 effect. Thus, it can be said that teachers' perceptions of situations and people are

541 significantly shaped by their expectations and biases is shaped (cf. Eckstein, 2018). How
542 strongly a disturbance is perceived and felt can also depend on psychological characteristics
543 such as self-efficacy beliefs (ibid. (ibid.). With decreasing self-efficacy, an increasing sense
544 of disturbance was found (ibid.).

545 The teacher-student relationship aspect, which includes recognizing student needs,
546 must also be considered. In order to be able to adequately counteract classroom
547 disruptions and thus automatically reduce the sense of disruption and the experience of
548 stress, it is important to address the concerns of the learners with empathy (Wickeren,
549 2019). Regarding gender as a factor, it should be noted that previous research has found
550 higher levels of disruptiveness in discipline problems among female teachers
551 (Kovess-Masfety, Rios-Seidel & Sevilla-Dedieu, 2007; cited in Wickeren, 2019, p. 33). In
552 summary can thus be summarized that the extent and perception of disturbances are based
553 on individual cognitive evaluation processes, the situational context, attentional focus,
554 didactic staging, and, indeed, professional experience (cf. Wickeren, 2019).

555 Research Questions and hypothesis

556 This study examined how the degree of teaching experience of novice and expert
557 teachers influences (1) the number of fixations on relevant areas, (2) the fixation duration
558 on relevant areas and (3) the time to first fixation on relevant events (e.g., the student
559 performing the disruption), using mobile eye-tracking data in a controlled, micro-teaching
560 setting. Furthermore, the study examined the individual evaluation of classroom
561 disruptions among experts and novices by assessing (4) how disruptive an event was and
562 (5) how confident the subjects felt dealing with the events. In addition, subjects' (6)
563 strategic knowledge of classroom management was assessed.

564 Based on the existing literature, we expected expert teachers to outperform novices
565 by (H1) showing more fixations on relevant areas with (H2) shorter fixation duration and

566 (H3) perceiving classroom disruptions faster (cf. van2014first?), (H4) feeling less troubled
567 by disruptions and (H5) more confident in dealing with disruptions (V. L. Barth, 2017),
568 and (H6) having higher scores in strategic knowledge of classroom management (Gold et
569 al., 2016).

570 **Materials and Methods**

571 We report how we determined our sample size, all data exclusions (if any), all
572 manipulations, and all measures in the study.

573 This study had a quasi-experimental study design, as there was no random
574 assignment of the subjects to experimental conditions. Due to the use of mobile
575 eye-tracking technology, the study had a high external validity (Gegenfurtner et al., 2018).
576 Internal validity could be ensured to the extent that the disruptive events that occurred
577 were exactly the same for all subjects, as the actors received precise behavioral
578 instructions. These disruptions followed a script and coding scheme in which the actions of
579 the class were precisely described. The order of the disruptions as well as the performing
580 actors were fully balanced with Latin Square to exclude row effects.

581 The scripted behavioral instructions during the micro-teaching unit characterized the
582 study with a high degree of standardization, especially when compared to events taking
583 place in a real classroom. The study was based on an experimental manual, script and
584 coding, which explicitly described the implementation, evaluation as well as interpretation
585 of the data, thus making it objectively recordable and measurable.

586 **Participants**

587 The sample consisted of $N = 84$ participants with $n = 42$ expert teachers and $n = 42$
588 novice teachers.

589 The inclusion criterion for experts was that they have successfully completed teacher

590 training and are actively employed in the teaching profession. According to Palmer,
591 Stough, Burdenski, and Gonzales (2005), we selected teachers as experts who had at least
592 three years of professional experience and ideally had worked in another teaching position,
593 such as subject advisor or trainer for trainee teachers, in addition to their teaching
594 profession in school. Novices were student teachers who had successfully completed their
595 first internship in a school and gained one to four hours of teaching experience.

596 The expert teachers (25 women; 59.52%) had a mean age of 39.30 years ($SD = 10.60$;
597 range: 26-60) and an average teaching experience of 11.80 years ($SD = 11.40$; range: 1-38).
598 9.52% of the experts were primary school teachers and 80.95% were secondary school
599 teachers. NA% of the experienced teachers were also engaged in an secondary teaching
600 activity, such as lecturers at the university, main training supervisors for trainee teachers
601 and subject advisers.

602 The novice teachers (29 women; 69.05%) had a mean age of 22.80 years ($SD = 1.90$;
603 range: 19-27) with an average teaching experience of 0 years. On average, the student
604 teachers were in their 6.70 semester ($SD = 2.60$; range: 3-11). Furthermore, they had an
605 average teaching experience of 9.60 teaching units à 45min ($SD = 7.20$; range: 0-36)
606 through the internships during their studies. 21.43% of the novices were studying to
607 become primary school teachers, 57.14% to become secondary school teachers and 19.05%
608 to become special education teachers. 88.10% of the student teachers were also engaged in
609 an extracurricular teaching activity, such as tutoring or homework supervision.

610 The subjects were primarily recruited through personal contacts, social media
611 (Facebook), e-mail distribution lists and advertising in lectures at the University Leipzig.
612 All study procedures were carried out in accordance with the ethical standards of the
613 University's Institutional Review Board. The authors received a positive vote on the study
614 procedures from the Ethics Committee Board of Leipzig University. All participants were
615 informed in detail about the aim and intention of the study prior to testing. Participation

616 in the study was voluntary and only took place after written consent has been given.

617 **Design**

618 The underlying research was conducted as a laboratory study in a cross-sectional
619 study design as part of the dissertation project *ProVisioNET*. This project investigates
620 whether and how a teacher's classroom experiences have an impact on the perception of
621 and response to classroom disruptions.

622 The study was based on a quasi-experimental study design, as there was no random
623 assignment of subjects to experimental conditions. Due to the use of MET technology, the
624 study had a high external validity (Gegenfurtner et al., 2018) To ensure internal validity
625 the occurring instructional events were exactly the same for all subjects, as the learners
626 received precise behavioral instructions. These disruptions followed a script and coding
627 guide in which the actions of the class were precisely described. The order of the
628 disruptions as well as the performing actors were prompted and fully balanced with Latin
629 Square. The scripted behavioral instructions during the classroom sequence distinguish
630 this study with a high degree of standardization, especially when compared to events that
631 take place in a real classroom. The study was based on an experimental manual, script,
632 and coding guide which explicitly describes the execution, evaluation and interpretation of
633 the data and thus and interpretation of the data, thus making them objectively
634 ascertainable and measurable.

635 **Materials and Measures**

636 **Eye-Tracking Data.**

637 **Apparatus.** To record eye-tracking data, teachers wore a binocular Tobii Pro
638 Glasses 2 eye-tracker during the micro-teaching-unit. The system consisted of a wearable
639 head unit and a recording unit. As shown in Figure 1, the head unit was a measuring

device with different sensors. A high-definition scene camera captured a full HD video of the teacher's field of vision. An integrated microphone recorded the surrounding sounds. Infrared light illuminators supported the eye tracking sensors which recorded the eye orientation to capture the teacher's gaze point as shown in Figure 2. The videos were recorded with a sampling rate of 50 Hz in a video resolution with 1920 x 1080 at 25 frames per second. The scene camera had a field of view of 90 deg. in 16:9 format (82 deg. horizontal and 52 deg. vertical) and a frame dimension of 179 x 159 x 57 mm (width x depth x height).

Teachers' visual focus of attention. Teachers' areas of interest (AOI) were identified as targets where teachers looked during the micro-teaching-unit using Tobii Pro Lab Analyzer Software Version 1.171. In the present study, general AOIs were defined as the student group, the students' nametags, the teacher material, the students' material, the board and non-instructional material in the classroom (wall, window, floor). These AOIs were coded during the manual fixation mapping process during the entire micro-teaching-unit. Additionally, the disruptive person was coded as an event-based AOI only over the duration of the disruption.

→ Only eye-tracking video recordings with 70 % and above gaze sample percentages were selected for the study to ensure that one or both eyes were detected during 70 % of the recording's duration (Chaudhuri et al, 2022). Screenshots of the video recordings were used to define the AOI in the eye-tracking videos. The coder started manual gaze mapping when the teacher entered the classroom and ended coding when the micro-teaching-unit ended. The start and end of the micro-teaching-unit was defined by a visual and acoustical signal. The two coders who were assigned to code the eye-tracking videos student employees who studied to be teachers and received a detailed introduction to the software. The intercoder reliability was checked by double coding 20 % of the videos from the whole dataset. Double coding agreement ranged from XXX % to XXX %.

We coded each fixation using the Attention Filter in Pro Lab with the velocity

667 threshold parameter set to 100 degrees/second instead of the default 30 degrees/second.
668 This filter is suitable for dynamic situations to study attention without separating any eye
669 movements. By choosing the Attention Filter typical fixations, most VOR (Vestibulo-ocular
670 reflex), smooth pursuit eye movements, and slow saccades are classified as fixations.

671 The relevant fixation metrics such as number of fixations, the total duration of
672 fixations and time to first fixation on the relevant event (disruptive person) were obtained
673 by exporting an interval-based TSV file from the Tobii Pro Lab Analyzer Software. The
674 total fixation duration and the time to first fixation metrics were measured in milliseconds
675 (ms) and the number of fixation in counts as an integer. To calculate the average fixation
676 of durations for all AOIs, we divided the total duration of fixations by the number of
677 fixations for each AOI.

678 **Video data.**

679 **Apparatus.** The speech, sounds and voices of the participants were recorded with
680 Zoom H3-VR Ambient Recorder installed in the middle of the lab setting (see set up plan
681 ???. The Zoom H3-VR recorded with four built-in mics arranged in an Ambisonic array
682 with a bitrate of 4608 kBits/s. Movements, facial expressions and gestures of the subjects
683 were recorded with four Go Pro Hero 7 black cameras from different angles (see set up plan
684 ???. The videos were recorded with a sampling rate of 50 Hz in a video resolution with 1920
685 x 1080 at 50 frames per second in 16:9 format with a linear field of view.

686 **Videography: Coding scheme for expertise indicators.** Part of the test
687 procedure was that the subject taught a 15min micro-teaching-unit and was disrupted by
688 the students in the class. The participant's behavior and reaction to the disruption were
689 recorded using video cameras and then evaluated with two self developed coding schemes
690 to assess the teacher's reaction on the one hand and the time on task during the unit. Both
691 aspects are important indicators of teaching quality and are therefore used as another
692 measure to make differences in expertise visible and measurable.

693 Video analysis was chosen as the research method for evaluating the teacher's
694 reaction to classroom disruptions because it offers numerous advantages. In contrast to
695 transcripts or audio-only recordings, a video also shows nonverbal forms of communication,
696 which is emphasized here, as well as the entire complex lesson excerpt in which many
697 things happen simultaneously (cf. Krammer & Reusser, 2012, p. 47f). Video data can also
698 be viewed and analyzed over and over again from different perspectives (Petko et al., 2003,
699 p. 265). The possible problem of selectivity (cf. Petko et al., 2003, p. 271), i.e., the fact
700 that only a section of the events can be shown due to the position and perspective of the
701 camera, was countered by recording the lessons from a total of four cameras. In this way, a
702 case of doubt can be analyzed from several angles. Another challenge can be the so-called
703 'camera effect', i.e. that the filmed persons feel disturbed and act differently than usual and
704 less naturally. However, Petko et al. (2003) conclude, despite filming with a camera person
705 at the time, that there is little reason to believe that teachers can easily change their
706 teaching style from one lesson to the next (p. 270) and that it should not be assumed that
707 the mere presence of a camera has significantly improved or worsened teaching' (p. 270). In
708 the 'ProVisioNET' survey, four very small cameras were placed in the room and their
709 presence was barely noticeable. Accordingly, it is assumed that no major nuisance
710 emanates from them.

711 Schnettler and Knoblauch (2009, p. 227) distinguish 'natural' and edited video data,
712 depending on whether the situation was created specifically for the recording or whether
713 the situation was influenced as little as possible or not at all. In the survey for this paper,
714 although the subjects were to act exactly as they would have done without a camera, the
715 classroom disruptions were scripted events that were performed by the actors.

716 *Teachers' Reaction to disruptions.* The evaluation of the teacher's reaction to the
717 disruptions was conducted as a systematic observation using a category system with fixed
718 categories and instructions. This corresponds to a highly inferential procedure, which is
719 combined with an event sampling procedure (cf. Lotz et al., 2013, p. 359).

720 A theory-based coding scheme was developed as an observation tool for classroom
721 observation. The coding of the reactions was done in seven levels. Based on the literature
722 on effective approaches to classroom disruptions, nonverbal response was given Code 1 as
723 the best variant because it best accommodates the classroom management strategies of
724 presence and overlapping (Gold & Holodynski, 2015; cf. Kounin, 2006) by not interrupting
725 the flow of instruction. Code 2 or Code 3 was assigned for a brief verbal expression with or
726 without a preceding nonverbal response. A brief, timely, and purposeful rebuke of the
727 disruptive student is considered by Kounin (2009, p. 99), among others, to be an important
728 component of the strategy of presence and thus demonstrates good classroom management.
729 If the teacher reacts non-verbally and at the same time with a longer verbal statement,
730 code 4 was assigned, because the flow of the lesson was already noticeably delayed. There
731 were two possibilities, if there was no reaction: If the teacher deliberately ignored the
732 disruption, code 5 was assigned. If the teacher did not notice the disruption, code 6 was
733 assigned, because then no effective monitoring was realized (cf. Gold & Holodynski,
734 2015). If the reaction to the disruption was dysfunctional, i.e., if the teacher shows an
735 inappropriate reaction, such as explaining the class rules for a very long time in the case of
736 an insignificant disruption, code 7 was assigned. According to Gippert et al. (2019), there
737 is a relational error in such cases, as the intensity of the student response should be relative
738 to the disruptive behavior (p. 2).

739 In developing the coding system, questions initially aroused about how descriptions of
740 responses should be made and in what order (especially codes 3-4). Initially, it was
741 planned to assign code 3 for a nonverbal and simultaneously long verbal response and code
742 4 for a purely verbal response. However, during the initial trial coding of the videos, it
743 became apparent that it was difficult to reliably distinguish nonverbal reactions to a
744 classroom disruption from normal facial expressions and gestures of the teacher when
745 talking. Moreover, during a prolonged verbal utterance, nonverbal signs could virtually
746 always be detected. Furthermore, a category for short, purely verbal reactions was missing.

747 After developing and reviewing the coding system, the observers coordinated and the
748 consistency of their ratings was verified. For the evaluation, the individual videos of the
749 sample were then coded and the data transferred to an Excel file.

750 Test quality criteria

751 In the construction of the video analysis method, attention was paid to compliance
752 with the quality criteria of objectivity, reliability and validity.

753 In order to ensure implementation objectivity, uniform implementation conditions
754 were created. In the experimental manual of the ProVisioNET study, the survey procedure
755 is described in detail, so that it is the same for each participant. On the one hand, the
756 objectivity of the evaluation was given by the processing in the R. On the other hand, prior
757 to the analysis of the videos, a precise coding scheme, which precisely describes the
758 individual codes, explains them and contains examples, was created and discursively
759 discussed in the team (see chapter 4.4.1). This coding scheme was tested and revised in a
760 piloting process, which increased the internal validity of the method. Interrater reliability
761 was determined for the two raters using sample ratings of the same videos. Percent
762 agreement yielded 77.8% and a Cohen's Kappa of $k = 0.6$, which is a satisfactory value
763 according to Landis & Koch (1977). Interpretive objectivity was ensured by specifying
764 scientific hypotheses that then had to be neutrally tested.

765 With the help of scripted behavioral instructions for students in the instructional
766 unit, a high degree of standardization was achieved that is not present in the normal school
767 setting. The lessons were recorded with several video cameras from different angles and an
768 audio recorder. This means a big advantage compared to the subjective evaluation of
769 observers in real time. In addition, video recordings can be viewed multiple times, by
770 different people or from different perspectives. Thus, a high measurement accuracy
771 (reliability) is given [cf. @mayring2005auswertung]. All measures taken contribute to
772 increase the validity of the method. However, it cannot be answered whether the coding

773 scheme is completely valid, because a validation with the help of another measuring
774 instrument was not possible. Therefore, it could not be controlled whether the collected
775 construct is related to other similar constructs.

776 *Coding Time on Task.* XXX

777 **Questionnaire data.**

778 **Presence Questionnaire.** After each micro-teaching-unit, the three actors
779 answered items on teaching quality using a validated questionnaire (Helmke et al., 2014)
780 and self developed scales on the teacher's presence behavior derived from the research
781 literature. In addition, subjects were asked to give a self-assessment on classroom
782 management by completing the same questionnaire after each micro-teaching-unit. The
783 questionnaire was a 4-point Likert scale (1 = Strongly Disagree; 2 = Disagree; 3 = Agree; 4
784 = Strongly Agree).

785 Bilanz: LB01_01 Bilanz: Die Lernenden haben in dieser Unterrichtslektion etwas
786 dazu gelernt. LB01_02 Bilanz: Die Lernenden haben sich in dieser Unterrichtslektion wohl
787 gefühlt. LB01_03 Bilanz: Mediennutzung und Sozialformen waren dem Inhalt und der
788 Situation der Unterrichtslektion angemessen.

789 Klassenmanagement: LM01_01 Klassenmanagement: Die gesamte Unterrichtslektion
790 wurde für den Lernstoff verwendet. LM01_02 Klassenmanagement: Ich habe alles
791 mitbekommen, was in der Klasse passiert ist. LM01_03 Klassenmanagement: Den
792 Lernenden war jederzeit klar, was sie tun sollten. LM01_04 Klassenmanagement: Die
793 Lernenden konnten ungestört arbeiten. LM01_05 Klassenmanagement: Die Lernenden
794 waren die ganze Unterrichtslektion über aktiv bei der Sache. LM01_06
795 Klassenmanagement: Ich habe vieles mit kurzen Blicken und knappen Gesten geregelt.
796 LM01_07 Klassenmanagement: Auf Störungen habe ich angemessen reagiert. LM01_08
797 Klassenmanagement: Bei Störungen gab ich den Lernenden ein klares STOP-Signal.

798 Redeanteil: LO01_01 Redeanteil: Wieviel Prozent der gesamten Unterrichtslektion

799 hat nach Ihrer Schätzung Ihr eigener Redeanteil eingenommen? ... % LO01_02

800 Redeanteil: Wenn ich Fragen oder Aufgaben gestellt habe, habe ich den Lernenden

801 folgende Zeit (in Sekunden) zum Nachdenken gegeben ... Sekunden

802 Präsenz: LP01_01 Präsenzindikatoren: Ich hatte eine aufrechte, der Klasse
803 zugewandte Körperhaltung. LP01_02 Präsenzindikatoren: Ich habe den Lernenden direkt
804 in die Augen geschaخت. LP01_03 Präsenzindikatoren: Ich habe meine Position im Raum
805 bewusst eingesetzt und variiert. LP01_04 Präsenzindikatoren: Ich habe alle oft angesehen.
806 LP01_05 Präsenzindikatoren: Meine Artikulation, Lautstärke und Tonhöhe waren
807 angemessen. LP01_06 Präsenzindikatoren: Meine Sprechgeschwindigkeit und
808 Sprechpausen waren angemessen. LP01_07 Präsenzindikatoren: Mein Sprechanteil an der
809 gesamten Sprechzeit der Unterrichtslektion war angemessen. LP01_08 Präsenzindikatoren:
810 Meine Gestik und Mimik waren abwechslungsreich und ausdrucksstark.

811 Natürliches Verhalten: LV01_01 Natürliches Verhalten: Während der

812 Unterrichtslektion habe ich mich sehr natürlich verhalten. LV01_02 Natürliches Verhalten:

813 Es war für mich kein Problem, vor einer fiktiven Klasse zu unterrichten. LV01_03

814 Natürliches Verhalten: Beim Unterrichten vor einer fiktiven Klasse habe ich mich so

815 verhalten, wie ich es auch in der Schule tun würde.

816 ***Teachers' perception of disruption and confidence using Rating scales.***

817 Two self-developed 11-point rating scales were used to record how disruptive the individual
818 disruptions were perceived by the subjects and how confident they felt in dealing with the
819 disruptions. Data for rating scales were collected during the Stimulted Recall Interview,
820 where the experimenter watched the pre-recorded eye-tracking video with the subject.

821 The questionnaire included nine questions for each scripted disruptive event. The

822 subjects did not have to complete the questionnaire themselves at any time, as this was

823 done by the experimenter.

824 For the present study, the following two questions from the interview were relevant:

825 1. "How disruptive was this event for you?" 2. "How confident did you feel in dealing with
826 this event?"

827 Both questions were answered using a scale ranging from 0 to 10. On this scale, 0
828 stood for "not at all disruptive" or "not at all safe" and 10 for "extremely disturbing" or
829 "extremely safe."

830 With these rating scales, the subjects indicated their personal subjective feeling
831 towards the individual lesson disruption. Rating scales usually consist of a number of
832 clearly arranged categories that can be presented visually in a variety of ways (Albers,
833 Klapper, Konradt, Walter, & Wolf, 2009, p. 67f). In the present questionnaire, the
834 numerical representation with eleven answer options (0-10) was used. There was no specific
835 designated neutral position. In this questionnaire, the middle category 5 was not
836 interpreted as a neutral position. In order to avoid misinterpretations of the scaling, the
837 questionnaire was described verbally by the experimenter and follow-up questions from the
838 test subjects were possible at any time. With regard to the quality criteria, it should be
839 made clear that complete objectivity in answering the two questions is given by the rating
840 scales. The question remained the same for all subjects and no additional comments were
841 made. Validity can also be assumed for the questionnaire, since the subjects could ask
842 follow-up questions and thus uncertainties were clarified. However, similar to the video
843 data, there was no perfect reliability with this survey instrument, because in a few
844 exceptional cases, for example, the experimenter forgot to ask the subject the relevant
845 questions.

846 ***Situational Jugdement Test of Strategic Knowledge of Classroom***

847 ***Management.*** To assess the strategic knowledge of classroom management, subjects
848 answered a Situational Judgement Test (SJT, (Gold & Holodynski, 2015)). In the
849 questionnaire, participants were given hypothetical school scenarios with five to six
850 response strategies for which they had to evaluate the effectiveness of each strategy on a
851 6-point rating scale from 1 (A) to 6 (F) according to school grades (see example scenario

852 XXX REFERENZ EINFÜGEN). Depending on the grade given to each strategy, scores
853 were converted in an SPSS syntax resulting in scores ranging from 0 to 1 for each
854 participant.

855 In developing the test procedure, the authors divided the concept of classroom
856 management into three main aspects: *Monitoring, managing momentum* (structuring) and
857 *rules and routines* (Gold & Holodynski, 2015, p. 230f).

858 *Monitoring* includes proactive strategies such as omnipresence. This means, the
859 teacher is always informed about the students' behavior, regulates it and reports it back to
860 them. The second strategy is overlapping. In this case, the teacher is able to control
861 several teaching processes at the same time. Monitoring means also reactive strategies of
862 supervision, which also includes the effective handling of classroom disruptions [cf. ibid.;
863 Kounin (2006), p. 89ff].

864 By *managing momentum* (structuring), Gold and Holodynski (2015) understand the
865 skillful control of teaching processes so that they can run smoothly and without delays,
866 taking into account the students' level of understanding and attention (Kounin, 2006, p.
867 101ff). They also include the group focus specifically identified by Kounin (Kounin, 2006,
868 p. 117ff), i.e. mobilizing the whole class and demanding accountability (Gold &
869 Holodynski, 2015, p. 231).

870 The third facet of classroom management, establishing *rules, routines and rituals*,
871 supports the other two facets and offers pupils orientation and structure in everyday school
872 life (cf. ibid.).

873 The construction of the SJT included 14 scenarios that were based on transcribed
874 classroom videos from Germany. For the 14 scenarios, response strategies were derived
875 from evidence-based and theoretical principles by 17 experts on classroom management.
876 With the exception of scenario 5, which was subsequently excluded, the 17 experts agreed
877 on an appropriate level of content validity for all other scenarios (cf. ibid.). For the

878 validation of the strategies, pair comparisons were made between different strategies and
879 the frequencies were calculated in relation to the expert responses (cf. ibid., p. 238).

880 Scenarios 10 and 12 were excluded from further analysis because they did not reach the
881 minimum level of agreement (cf. ibid.).

882 The construct validity of the SJT with the other 11 scenarios and the sensitivity to
883 differences in strategic knowledge was then examined in a pilot study with 98 trainee
884 teachers and their results validated in a cross-validation with a larger sample (cf. ibid.,
885 p. 238ff). Despite limitations in reliability, the test was confirmed to have content and
886 construct validity (cf. ibid., p. 243). On the basis of these results, sufficient reliability and
887 validity of the SJT test procedure was assumed.

888 The scenarios could each be assigned to one facet of the mentioned classroom
889 management: Scenarios 1-4 cover the facet *monitoring*, scenarios 6-9 the facet *managing*
890 *momentum* and scenarios 11-14 the facet *rules and routines*.

891 In the current study, objectivity of application was ensured by the
892 questionnaire-format, the written instructions and by a detailed test administration
893 manual of the study *ProVisioNET*. The objectivity of analysis was given by processing the
894 obtained questionnaire data in R (Version 4.3.3; R Core Team, 2021) and the R-packages
895 *ARTofR* (Version 0.4.1; Zhang, 2021), *cowplot* (Version 1.1.3; Wilke, 2020), *DescTools*
896 (Version 0.99.54; Andri et mult. al., 2022), *dplyr* (Version 1.1.4; Wickham, François, Henry,
897 & Müller, 2022), *forcats* (Version 1.0.0; Wickham, 2021), *formattable* (Version 0.2.1; Ren &
898 Russell, 2021), *ggmosaic* (Jeppson, Hofmann, & Cook, 2021), *ggplot2* (Version 3.5.1;
899 Wickham, 2016), *ggrepel* (Version 0.9.5; Slowikowski, 2021), *ggthemes* (Version 5.1.0;
900 Arnold, 2021), *gridExtra* (Version 2.3; Auguie, 2017), *haven* (Version 2.5.4; Wickham &
901 Miller, 2021), *imputeTS* (Version 3.3; Moritz & Bartz-Beielstein, 2017), *janitor* (Version
902 2.2.0; Firke, 2021), *kableExtra* (Version 1.4.0; Zhu, 2021), *knitr* (Version 1.47; Xie, 2015),
903 *lme4* (Version 1.1.35.4; Bates, Mächler, Bolker, & Walker, 2015), *ltm* (Version 1.2.0;

904 Rizopoulos, 2006), *lubridate* (Version 1.9.3; Grolemund & Wickham, 2011), *MASS* (Version
905 7.3.60.0.1; Venables & Ripley, 2002), *Matrix* (Version 1.6.5; Bates & Maechler, 2021),
906 *moments* (Version 0.14.1; Komsta & Novomestky, 2015), *msm* (Version 1.7.1; Jackson,
907 2011), *needs* (Version 0.0.3; Katz, 2016), *papaja* (Version 0.1.2; Aust & Barth, 2020),
908 *polycor* (Version 0.8.1; Fox, 2022), *purrr* (Version 1.0.2; Henry & Wickham, 2020), *readr*
909 (Version 2.1.5; Wickham, Hester, & Bryan, 2021), *readxl* (Version 1.4.3; Wickham &
910 Bryan, 2019), *rgeos* (Bivand & Rundel, 2021), *rlang* (Version 1.1.4; Henry & Wickham,
911 2022), *rnaturrearth* (South, 2017a, 2017b), *rnaturrearthdata* (South, 2017b), *sf* (E.
912 Pebesma, 2018), *sjlabelled* (Version 1.2.0; Lüdecke, 2022), *sjPlot* (Version 2.8.16; Lüdecke,
913 2021), *sp* (E. J. Pebesma & Bivand, 2005), *stringr* (Version 1.5.1; Wickham, 2019), *tibble*
914 (Version 3.2.1; Müller & Wickham, 2021), *tidyverse* (Version 2.0.0; Wickham et al., 2019), *tinylabels* (Version 0.2.4; M. Barth, 2022),
915 *viridis* (Version 0.6.5; Garnier et al., 2021a, 2021b), *viridisLite* (Version 0.4.2; Garnier et
916 al., 2021b), and *xtable* (Version 1.8.4; Dahl, Scott, Roosen, Magnusson, & Swinton, 2019)
917 and IBM SPSS 28.

919 Since the test was a questionnaire designed to assess strategic knowledge of classroom
920 management of *primary school teachers*, but the sample of the study included several types
921 of schools, changes were made in 10 out of 12 scenarios by deleting the indicated class
922 grade. Only in two scenarios was the class grade information not deleted, as the action
923 alternatives would otherwise not be appropriate without the information.

924 XXX FRANZI

925 Procedure

926 The project was conducted as a laboratory study in a cross-sectional study design to
927 investigate whether and how teachers' experience has an influence on the perception of and
928 reaction to classroom disruptions.

929 In June 2021, the study was piloted with student teachers volunteers to refine the
930 study procedure. Data collection was conducted between July 2021 and Septemper 2022.

931 Before the data collection, each subject received a personal digital meeting with the
932 experimenter to go over the study procedure and to arrange a date for the data collection.
933 During the digital meeting, the subjects were asked to prepare a 15-minute lesson in a
934 subject and grade of their choice for the data collection.

935 On the day of a data collection, the first step was to set up the study room at the
936 University of Leipzig. For this, all the appropriate technical equipment was charged and
937 installed in the room (see set up plan [REFERENZ EINFÜGEN]). Next, all four cameras
938 and the audio recorder were synced via Timecode System.

939 After the subject arrived, a smart watch was put on to measure the heart rate during
940 the session and to get a pretest time at least 15min before the session started. In addition
941 to the experimenter and the subject, three student assistants from the working group
942 always took part in the data collection, as they represented the class.

943 After the welcome, the subject was again briefed about the study. It was explicitly
944 pointed out that the student assistants would act as the class and simulate typical class
945 events during the lesson. The subject was asked in advance to behave as naturally as
946 possible during the entire time. Next, the subjects' written informed consent was obtained
947 and contact details were collected in order to inform all persons participating in the study
948 if a covid infection should occur.

949 After the introduction, the eye-tracking glasses were adjusted for the subject by
950 inserting contact lenses if necessary and changing the nose pad. To start the eye tracking
951 glasses, the Tobii Glasses must be fitted onto the subject's head via an
952 one-point-calibration. In the calibration process the subject was asked to look at a
953 Calibration Card held in-front of the subject for a few seconds. The experimenter started
954 the recording from Tobii Glasses Controller Software running on a computer.

955 After starting the eye tracking recording, all other technical devices were also
956 switched on: The four cameras and the audio recorder were controlled via iPad using the
957 BLINK Hup app and could be started simultaneously by synchronization. The ZED
958 camera was started manually on another laptop.

959 Before the 15-minute lesson, there was a short 10-minute warm-up phase. The phase
960 was divided into two parts and served on the one hand to get the subjects used to the
961 eye-tracking glasses and on the other hand to get used to their class. In the first phase of
962 the warm-up, the game “name juggling” was played using two balls. In the game, the
963 subject and the three actors threw two balls at each other and, depending on the color of
964 the ball, called either their own name or that of the target person. After the name
965 juggling, the subjects were supposed to start a conversation. For this, the subject thought
966 of a question for each student and was also asked a question of each student. The content
967 could be anything that interested the participants.

968 After the warm up phase, the experimenter ensured a manual synchronization of the
969 technique by an acoustic signal in which she clapped her hands loudly twice standing in the
970 middle of the room. After this, another nine-point calibration followed outside the study
971 room in a neighboring room. Before the subject left the room for calibration, the time on
972 the smartwatch was noted, as well as the steps recorded until that point. The subject had
973 to stand at a marked point and look at a board three meters away with nine april tags.
974 The subject was asked to read the nine points aloud in order at a normal speaking speed.
975 This procedure was important to validate the one-point calibration on the one hand and on
976 the other hand to give the subject the feeling of a lesson start, because after this
977 calibration the subject came into the study room to start the 15-minute lesson.

978 For the micro-teaching lesson, student teachers and experienced teachers were asked
979 to prepare an introduction of 15 minutes which they taught in front of the fictitious class
980 consisting of three student assistants. The actors simulated the nine classroom events

981 during the lesson, derived by research literature. The order of the disruptions as well as the
982 students performing them were fully balanced using Latin Square. The disruptions were
983 only visible for the class on a screen.

984 During the lesson, a mobile eye-tracker recorded the subject's gaze behavior and
985 audio data of the lesson. All other sounds and voices were recorded by an audio recorder.
986 To record facial expressions, gestures and movements, four mobile cameras were installed
987 to record the classroom from all perspectives (!!see figures).

988 After the lesson, the time was again noted from the smartwatch as well as the
989 subject's steps. The nine-point calibration was also performed again in the neighboring
990 room. This time, however, the subject was asked to wait outside the room until he or she
991 was called in, because four letters from A to D were placed in the study room. The subject
992 was asked to stand facing the board at a marked point in the room and, when given an
993 acoustic signal, to turn around and search the letters and read them aloud in order. This
994 served as a control condition for the speed of the subjects' perceptual ability as no
995 expertise is required for searching letters.

996 After the letter search, the experimenter clapped twice with the hands to record an
997 acoustic signal for the synchronization of all technical devices. Afterwards, all devices were
998 switched off and the subjects as well as the actors were given a short questionnaire, which
999 contained items to collect demographic information as well as items about the previously
1000 given lesson on teaching quality using a validated questionnaire (Helmke et al., 2014) and
1001 self developed scales on the teacher's presence behavior derived from the research
1002 literature. The completion of the questionnaire took approximately 5 minutes.

1003 In the second part of the study, the experimenter conducted a Stimulated Recall
1004 Interview (SRI) with the subject(see in Figure 3. The interview was recorded by the
1005 ambient recorder and additionally with a software called OBS to record the speech and
1006 screen. The recorded eye-tracking video was watched and commented on by the subject

1007 while thinking aloud. The interview was structured by a guideline based on a fixed
1008 sequence of questions, however the subjects were allowed to answer completely open to
1009 these questions. The objective was to collect data that would provide insight into the
1010 subjects' thoughts and reactions in relation to the scripted classroom disruptions. The
1011 eye-tracking video was not completely re-watched, but stopped at the points where
1012 disruptions occurred. After the disruption occurred, the subject was asked to describe
1013 what had just happened. If the disruption was noticed, the subject was asked how
1014 disruptive the event was at that moment on a scale from 0 to 10. In addition, the subject
1015 was asked to reason the given number. In a next step, the experimenter re-played the
1016 subject's reaction to the disruption and asked to describe and reason the reaction. The
1017 experimenter then asked the subject to rate on a scale from 0 to 10 how confident he/she
1018 felt in dealing with the disruption and again to reason the number (for a detailed
1019 description of the scale question, see Rating Scales).

1020 Finally, the subjects answered a Situational Judgement Test of Strategic Knowledge
1021 of Classroom Management (SJT, Gold & Holodynski, 2015) in the form of a questionnaire.
1022 Here they had to assess teaching scenarios and evaluate their behavior in response to them.
1023 The SJT was used to assess strategic knowledge about classroom management. The
1024 completion of the entire survey took 15 min on average.

1025 Data analysis

1026 We investigated whether experts and novice teachers differed

1027 All reported data analyses were conducted with the R (Version 4.3.3; R Core Team,
1028 2021) and the R-packages *ARTofR* (Version 0.4.1; Zhang, 2021), *cowplot* (Version 1.1.3;
1029 Wilke, 2020), *DescTools* (Version 0.99.54; Andri et mult. al., 2022), *dplyr* (Version 1.1.4;
1030 Wickham et al., 2022), *forcats* (Version 1.0.0; Wickham, 2021), *formattable* (Version 0.2.1;
1031 Ren & Russell, 2021), *ggmosaic* (Jeppson et al., 2021), *ggplot2* (Version 3.5.1; Wickham,

1032 2016), *ggrepel* (Version 0.9.5; Slowikowski, 2021), *ggthemes* (Version 5.1.0; Arnold, 2021),
1033 *gridExtra* (Version 2.3; Auguie, 2017), *haven* (Version 2.5.4; Wickham & Miller, 2021),
1034 *imputeTS* (Version 3.3; Moritz & Bartz-Beielstein, 2017), *janitor* (Version 2.2.0; Firke,
1035 2021), *kableExtra* (Version 1.4.0; Zhu, 2021), *knitr* (Version 1.47; Xie, 2015), *lme4* (Version
1036 1.1.35.4; Bates et al., 2015), *ltm* (Version 1.2.0; Rizopoulos, 2006), *lubridate* (Version 1.9.3;
1037 Grolemund & Wickham, 2011), *MASS* (Version 7.3.60.0.1; Venables & Ripley, 2002),
1038 *Matrix* (Version 1.6.5; Bates & Maechler, 2021), *moments* (Version 0.14.1; Komsta &
1039 Novomestky, 2015), *msm* (Version 1.7.1; Jackson, 2011), *needs* (Version 0.0.3; Katz, 2016),
1040 *papaja* (Version 0.1.2; Aust & Barth, 2020), *polycor* (Version 0.8.1; Fox, 2022), *purrr*
1041 (Version 1.0.2; Henry & Wickham, 2020), *readr* (Version 2.1.5; Wickham et al., 2021),
1042 *readxl* (Version 1.4.3; Wickham & Bryan, 2019), *rgeos* (Bivand & Rundel, 2021), *rlang*
1043 (Version 1.1.4; Henry & Wickham, 2022), *rnaturalearth* (South, 2017a, 2017b),
1044 *rnaturalearthdata* (South, 2017b), *sf* (E. Pebesma, 2018), *sjlabelled* (Version 1.2.0; Lüdecke,
1045 2022), *sjPlot* (Version 2.8.16; Lüdecke, 2021), *sp* (E. J. Pebesma & Bivand, 2005), *stringr*
1046 (Version 1.5.1; Wickham, 2019), *tibble* (Version 3.2.1; Müller & Wickham, 2021), *tidyverse*
1047 (Version 1.3.1; Wickham & Girlich, 2022), *tidyverse* (Version 2.0.0; Wickham et al., 2019),
1048 *tinylabels* (Version 0.2.4; M. Barth, 2022), *viridis* (Version 0.6.5; Garnier et al., 2021a,
1049 2021b), *viridisLite* (Version 0.4.2; Garnier et al., 2021b), and *xtable* (Version 1.8.4; Dahl et
1050 al., 2019) and IBM SPSS 28.

1051

Results

1052

Discussion

1053

References

- 1054 Albers, S., Klapper, D., Konradt, U., Walter, A., & Wolf, J. (2009). *Methodik der empirischen forschung* (Vol. 3). Springer.
- 1055
- 1056 Andri et mult. al., S. (2022). *DescTools: Tools for descriptive statistics*. Retrieved from
1057 <https://cran.r-project.org/package=DescTools>
- 1058 Arnold, J. B. (2021). *Ggthemes: Extra themes, scales and geoms for 'ggplot2'*. Retrieved
1059 from <https://CRAN.R-project.org/package=ggthemes>
- 1060 Auguie, B. (2017). *gridExtra: Miscellaneous functions for "grid" graphics*. Retrieved from
1061 <https://CRAN.R-project.org/package=gridExtra>
- 1062 Aust, F., & Barth, M. (2020). *papaja: Prepare reproducible APA journal articles with R
1063 Markdown*. Retrieved from <https://github.com/crsh/papaja>
- 1064 Barth, M. (2022). *tinylabes: Lightweight variable labels*. Retrieved from
1065 <https://cran.r-project.org/package=tinylabes>
- 1066 Barth, V. L. (2017). *Professionelle wahrnehmung von störungen im unterricht*. Springer.
- 1067 Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects
1068 models using lme4. *Journal of Statistical Software*, 67(1), 1–48.
1069 <https://doi.org/10.18637/jss.v067.i01>
- 1070 Bates, D., & Maechler, M. (2021). *Matrix: Sparse and dense matrix classes and methods*.
1071 Retrieved from <https://CRAN.R-project.org/package=Matrix>
- 1072 Baumert, J., & Kunter, M. (2006). Key word: Professional competence of teachers.
1073 *ZEITSCHRIFT FUR ERZIEHUNGSWISSENSCHAFT*, 9(4), 469–520.
- 1074 Berliner, D. C. (1994). Expertise: The wonders of exemplary performance. *Creating
1075 Powerful Thinking in Teachers and Students*, 161–186.
- 1076 Bivand, R., & Rundel, C. (2021). *Rgeos: Interface to geometry engine - open source
1077 ('GEOS')*. Retrieved from <https://CRAN.R-project.org/package=rgeos>
- 1078 Blake, C. (2013). Eye-tracking: Grundlagen und anwendungsfelder. In *Handbuch
1079 standardisierte erhebungsverfahren in der kommunikationswissenschaft* (pp. 367–387).

- 1080 Springer.
- 1081 Dahl, D. B., Scott, D., Roosen, C., Magnusson, A., & Swinton, J. (2019). *Xtable: Export*
1082 *tables to LaTeX or HTML*. Retrieved from
1083 <https://CRAN.R-project.org/package=xtable>
- 1084 Eckstein, B., Grob, U., & Reusser, K. (2016). Unterrichtliche devianz und subjektives
1085 störungsempfinden. Entwicklung eines instrumentariums zur erfassung von
1086 unterrichtsstörungen. *Empirische Pädagogik (EP)*, 30(1), 113–129.
- 1087 Emmer, E., Sanford, J., Clements, B., & Martin, J. (1982). Improving classroom
1088 management and organization in junior high schools. *An. Experimental, Investigation.*
1089 *Washington, DC: National Institute*.
- 1090 Epstein, R. M., & Hundert, E. M. (2002). Defining and assessing professional competence.
1091 *Jama*, 287(2), 226–235.
- 1092 Evertson, C. M., Weinstein, C. S., et al. (2006). Classroom management as a field of
1093 inquiry. *Handbook of Classroom Management: Research, Practice, and Contemporary*
1094 *Issues*, 3(1), 16.
- 1095 Firke, S. (2021). *Janitor: Simple tools for examining and cleaning dirty data*. Retrieved
1096 from <https://CRAN.R-project.org/package=janitor>
- 1097 Fox, J. (2022). *Polycor: Polychoric and polyserial correlations*. Retrieved from
1098 <https://CRAN.R-project.org/package=polycor>
- 1099 Garnier, Simon, Ross, Noam, Rudis, Robert, ... Cédric. (2021a). *viridis -*
1100 *colorblind-friendly color maps for r*. <https://doi.org/10.5281/zenodo.4679424>
- 1101 Garnier, Simon, Ross, Noam, Rudis, Robert, ... Cédric. (2021b). *viridis -*
1102 *colorblind-friendly color maps for r*. <https://doi.org/10.5281/zenodo.4679424>
- 1103 Gegenfurtner, A., Eichinger, A., Latzel, R., Dietrich, M. P., Barkowsky, M., Glufke, A., ...
1104 Stern, W. (2018). *Mobiles eye-tracking in den angewandten wissenschaften*.
- 1105 Gold, B., & Holodynski, M. (2015). Development and construct validation of a situational
1106 judgment test of strategic knowledge of classroom management in elementary schools.

- 1107 *Educational Assessment*, 20(3), 226–248.
- 1108 Gold, B., & Holodynski, M. (2017). Using digital video to measure the professional vision
1109 of elementary classroom management: Test validation and methodological challenges.
1110 *Computers & Education*, 107, 13–30.
- 1111 Gromlund, G., & Wickham, H. (2011). Dates and times made easy with lubridate.
1112 *Journal of Statistical Software*, 40(3), 1–25. Retrieved from
1113 <https://www.jstatsoft.org/v40/i03/>
- 1114 Grub, A.-S., Biermann, A., & Brünken, R. (2020). Process-based measurement of
1115 professional vision of (prospective) teachers in the field of classroom management. A
1116 systematic review. *Journal for Educational Research Online*, 12(3), 75–102.
- 1117 Helmke, A., Helmke, T., Lenske, G., Pham, G., Praetorius, A.-K., Schrader, F.-W., &
1118 AdeThurow, M. (2014). Unterrichtsdiagnostik mit EMU. *Aus- Und Fortbildung Der
1119 Lehrkräfte in Hinblick Auf Verbesserung Der Diagnosefähigkeit, Umgang Mit
1120 Heterogenität Und Individuelle Förderung*, 149–163.
- 1121 Henry, L., & Wickham, H. (2020). *Purrr: Functional programming tools*. Retrieved from
1122 <https://CRAN.R-project.org/package=purrr>
- 1123 Henry, L., & Wickham, H. (2022). *Rlang: Functions for base types and core r and
1124 'tidyverse' features*. Retrieved from <https://CRAN.R-project.org/package=rlang>
- 1125 Jackson, C. H. (2011). Multi-state models for panel data: The msm package for R. *Journal
1126 of Statistical Software*, 38(8), 1–29. Retrieved from <https://www.jstatsoft.org/v38/i08/>
- 1127 Jahn, G., Stürmer, K., Seidel, T., & Prenzel, M. (2014). Professionelle
1128 unterrichtswahrnehmung von lehramtsstudierenden: Eine scaling-up studie des
1129 observe-projekts. *Zeitschrift für Entwicklungspsychologie Und pädagogische Psychologie*,
1130 46(4), 171–180.
- 1131 Jeppson, H., Hofmann, H., & Cook, D. (2021). *Ggmosaic: Mosaic plots in the 'ggplot2'
1132 framework*. Retrieved from <https://CRAN.R-project.org/package=ggmosaic>
- 1133 Katz, J. (2016). *Needs: Attaches and installs packages*. Retrieved from

- 1134 https://CRAN.R-project.org/package=needs
- 1135 Kleinknecht, M., & Schneider, J. (2013). What do teachers think and feel when analyzing
1136 videos of themselves and other teachers teaching? *Teaching and Teacher Education*, 33,
1137 13–23.
- 1138 Klieme, E., Hartig, J., & Rauch, D. (2008). The concept of competence in educational
1139 contexts. *Assessment of Competencies in Educational Contexts*, 3, 22.
- 1140 Komsta, L., & Novomestky, F. (2015). *Moments: Moments, cumulants, skewness, kurtosis
1141 and related tests*. Retrieved from https://CRAN.R-project.org/package=moments
- 1142 Kounin, J. S. (2006). *Techniken der klassenführung*. Waxmann Verlag.
- 1143 Lachner, A., Jarodzka, H., & Nückles, M. (2016). What makes an expert teacher?
1144 Investigating teachers' professional vision and discourse abilities. *Instructional Science*,
1145 44(3), 197–203.
- 1146 Leinhardt, G., & Greeno, J. G. (1986). The cognitive skill of teaching. *Journal of
1147 Educational Psychology*, 78(2), 75.
- 1148 Lüdecke, D. (2021). *sjPlot: Data visualization for statistics in social science*. Retrieved
1149 from https://CRAN.R-project.org/package=sjPlot
- 1150 Lüdecke, D. (2022). *Sjlabelled: Labelled data utility functions (version 1.2.0)*.
1151 https://doi.org/10.5281/zenodo.1249215
- 1152 Marzano, R. J., Marzano, J. S., & Pickering, D. (2003). *Classroom management that
1153 works: Research-based strategies for every teacher*. ASCD.
- 1154 Moritz, S., & Bartz-Beielstein, T. (2017). imputeTS: Time Series Missing Value
1155 Imputation in R. *The R Journal*, 9(1), 207–218. https://doi.org/10.32614/RJ-2017-009
- 1156 Müller, K., & Wickham, H. (2021). *Tibble: Simple data frames*. Retrieved from
1157 https://CRAN.R-project.org/package=tibble
- 1158 Oser, F. K., Achtenhagen, F., & Renold, U. (2006). Competence-oriented teacher training:
1159 Old research demands and new pathways. In *Competence oriented teacher training* (pp.
1160 1–7). Brill.

- 1161 Palmer, D. J., Stough, L. M., Burdenski, T. K., Jr, & Gonzales, M. (2005). Identifying
1162 teacher expertise: An examination of researchers' decision making. *Educational
1163 Psychologist, 40*(1), 13–25.
- 1164 Pebesma, E. (2018). Simple Features for R: Standardized Support for Spatial Vector Data.
1165 *The R Journal, 10*(1), 439–446. <https://doi.org/10.32614/RJ-2018-009>
- 1166 Pebesma, E. J., & Bivand, R. S. (2005). Classes and methods for spatial data in R. *R
1167 News, 5*(2), 9–13. Retrieved from <https://CRAN.R-project.org/doc/Rnews/>
- 1168 R Core Team. (2021). *R: A language and environment for statistical computing*. Vienna,
1169 Austria: R Foundation for Statistical Computing. Retrieved from
1170 <https://www.R-project.org/>
- 1171 Ren, K., & Russell, K. (2021). *Formattable: Create 'formattable' data structures*. Retrieved
1172 from <https://CRAN.R-project.org/package=formattable>
- 1173 Rizopoulos, D. (2006). Ltm: An r package for latent variable modelling and item response
1174 theory analyses. *Journal of Statistical Software, 17*(5), 1–25. Retrieved from
1175 <https://doi.org/10.18637/jss.v017.i05>
- 1176 Sabers, D. S., Cushing, K. S., & Berliner, D. C. (1991). Differences among teachers in a
1177 task characterized by simultaneity, multidimensional, and immediacy. *American
1178 Educational Research Journal, 28*(1), 63–88.
- 1179 Seidel, T., & Stürmer, K. (2014). Modeling and measuring the structure of professional
1180 vision in preservice teachers. *American Educational Research Journal, 51*(4), 739–771.
- 1181 Sherin, M. (2007). *The development of teachers' professional vision in video clubs. Video
1182 research in the learning sciences*. R. Goldman, r. Pea, b. Barron and SJ derry.
1183 Mahwah, NJ, Lawrence Erlbaum.
- 1184 Slowikowski, K. (2021). *Ggrepel: Automatically position non-overlapping text labels with
1185 'ggplot2'*. Retrieved from <https://CRAN.R-project.org/package=ggrepel>
- 1186 South, A. (2017a). *Rnaturalearth: World map data from natural earth*. Retrieved from
1187 <https://CRAN.R-project.org/package=rnaturalearth>

- 1188 South, A. (2017b). *Rnaturalearthdata: World vector map data from natural earth used in*
1189 '*rnatueralearth*'. Retrieved from
1190 <https://CRAN.R-project.org/package=rnaturalearthdata>
- 1191 Treisch, F. (2018). *Die entwicklung der professionellen unterrichtswahrnehmung im*
1192 *lehr-lern-labor seminar*. Bayerische Julius-Maximilians-Universitaet Wuerzburg
1193 (Germany).
- 1194 Venables, W. N., & Ripley, B. D. (2002). *Modern applied statistics with s* (Fourth). New
1195 York: Springer. Retrieved from <https://www.stats.ox.ac.uk/pub/MASS4/>
- 1196 Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Springer-Verlag New
1197 York. Retrieved from <https://ggplot2.tidyverse.org>
- 1198 Wickham, H. (2019). *Stringr: Simple, consistent wrappers for common string operations*.
1199 Retrieved from <https://CRAN.R-project.org/package=stringr>
- 1200 Wickham, H. (2021). *Forcats: Tools for working with categorical variables (factors)*.
1201 Retrieved from <https://CRAN.R-project.org/package=forcats>
- 1202 Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., ...
1203 Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, 4(43),
1204 1686. <https://doi.org/10.21105/joss.01686>
- 1205 Wickham, H., & Bryan, J. (2019). *Readxl: Read excel files*. Retrieved from
1206 <https://CRAN.R-project.org/package=readxl>
- 1207 Wickham, H., François, R., Henry, L., & Müller, K. (2022). *Dplyr: A grammar of data*
1208 *manipulation*. Retrieved from <https://CRAN.R-project.org/package=dplyr>
- 1209 Wickham, H., & Girlich, M. (2022). *Tidyr: Tidy messy data*. Retrieved from
1210 <https://CRAN.R-project.org/package=tidyr>
- 1211 Wickham, H., Hester, J., & Bryan, J. (2021). *Readr: Read rectangular text data*. Retrieved
1212 from <https://CRAN.R-project.org/package=readr>
- 1213 Wickham, H., & Miller, E. (2021). *Haven: Import and export 'SPSS', 'stata' and 'SAS'*
1214 *files*. Retrieved from <https://CRAN.R-project.org/package=haven>

- 1215 Wilke, C. O. (2020). *Cowplot: Streamlined plot theme and plot annotations for 'ggplot2'*.
1216 Retrieved from <https://CRAN.R-project.org/package=cowplot>
- 1217 Wolff, C. E., Bogert, N. van den, Jarodzka, H., & Boshuizen, H. P. (2015). Keeping an eye
1218 on learning: Differences between expert and novice teachers' representations of
1219 classroom management events. *Journal of Teacher Education*, 66(1), 68–85.
- 1220 Wolff, C. E., Jarodzka, H., Bogert, N. van den, & Boshuizen, H. (2016). Teacher vision:
1221 Expert and novice teachers' perception of problematic classroom management scenes.
1222 *Instructional Science*, 44(3), 243–265.
- 1223 Xie, Y. (2015). *Dynamic documents with R and knitr* (2nd ed.). Boca Raton, Florida:
1224 Chapman; Hall/CRC. Retrieved from <https://yihui.org/knitr/>
- 1225 Zhang, H. (2021). *ARTofR: Who ever care about the [art of r] scripts?* Retrieved from
1226 <https://CRAN.R-project.org/package=ARTofR>
- 1227 Zhu, H. (2021). *kableExtra: Construct complex table with 'kable' and pipe syntax*.
1228 Retrieved from <https://CRAN.R-project.org/package=kableExtra>

Table 1

Demographic Information

Group	N	Gender female in percent	M Age in years	SD Age in years	Min Age in years	Max Age in years
Expert	42	59.52	39.31	10.64	26.00	60.00
Novice	42	69.05	22.83	1.85	19.00	27.00

Table 2

Teaching Experience in years, internship experience in teaching units (45min) and extracurricular teaching units (45min)

Group	N	M Exp.	SD Exp.	Min Exp.	Max Exp.	M Semester	SD Semester	Min Semester	Max Semester	M Internship	SD Internship
Expert	42	11.79	11.43	1.00	38.00	NA	NA	NA	NA	NA	NA
Novice	42	0.00	0.00	0.00	0.00	6.74	2.65	3.00	11.00	9.62	7.19

Table 3

Scale analysis for teachers' self-assessment

	N Items	M	SD	Min	Max	Skewness	Kurtosis	Alpha
Classroom Management	8.00	3.09	0.35	2.38	3.88	0.29	2.45	0.60
Balance	3.00	3.27	0.52	1.67	4.00	-0.49	3.14	0.69
Presence	8.00	3.21	0.40	2.12	3.88	-0.38	2.44	0.74
Natural Behavior	3.00	3.25	0.63	1.33	4.00	-0.72	3.14	0.78

Table 4

Table Rating Scales Disruption Factor

Group	Event	M	SD	Min	Max
Expert	chatting	6.44	3.33	1.00	10.00
Expert	clicking pen	5.57	2.87	0.00	10.00
Expert	drawing	1.85	1.46	0.00	6.00
Expert	drumming	4.60	2.44	1.00	8.00
Expert	head on table	3.73	2.25	0.00	8.00
Expert	heckling	7.19	2.37	2.00	10.00
Expert	looking at phone	4.71	3.10	1.00	10.00
Expert	snipping	4.00	3.61	0.00	10.00
Expert	whispering	4.56	2.50	0.00	9.00
Novice	chatting	8.62	1.35	5.00	10.00
Novice	clicking pen	7.00	2.11	3.00	10.00
Novice	drawing	2.24	1.48	0.00	4.00
Novice	drumming	6.57	2.27	0.00	10.00
Novice	head on table	4.09	2.09	1.00	8.00
Novice	heckling	6.96	2.57	2.00	10.00
Novice	looking at phone	3.90	1.94	1.00	8.00
Novice	snipping	4.67	3.33	0.00	9.00
Novice	whispering	6.08	2.36	1.00	10.00

Table 5

Table Rating Scales Confident Factor

Group	Event	M	SD	Min	Max
Expert	chatting	8.44	2.03	2.00	10.00
Expert	clicking pen	8.93	1.00	7.00	10.00
Expert	drawing	9.23	1.36	5.00	10.00
Expert	drumming	9.07	0.96	7.00	10.00
Expert	head on table	9.00	1.41	5.00	10.00
Expert	heckling	6.75	2.89	2.00	10.00
Expert	looking at phone	9.21	1.05	7.00	10.00
Expert	snipping	9.06	1.48	4.00	10.00
Expert	whispering	8.06	2.11	3.00	10.00
Novice	chatting	6.42	2.39	0.00	10.00
Novice	clicking pen	7.54	1.89	3.00	10.00
Novice	drawing	8.81	1.25	5.00	10.00
Novice	drumming	7.57	2.11	3.00	10.00
Novice	head on table	7.30	1.96	3.00	10.00
Novice	heckling	5.35	2.53	1.00	10.00
Novice	looking at phone	7.65	1.73	3.00	10.00
Novice	snipping	8.04	1.73	3.00	10.00
Novice	whispering	7.25	2.03	2.00	10.00

Table 6

Scale analysis SJT novices

	N Items	M	SD	Min	Max	Skewness	Kurtosis	alpha
Monitoring	4.00	0.71	0.11	0.44	0.96	-0.11	3.76	0.37
Managing Momentum	5.00	0.74	0.11	0.51	0.94	-0.25	2.85	0.19
Rules and routines	3.00	0.75	0.11	0.46	0.90	-0.74	3.25	-0.28

Table 7

Scale analysis SJT novices

	N Items	M	SD	Min	Max	Skewness	Kurtosis	alpha
Monitoring	4.00	0.71	0.15	0.52	0.92	0.17	1.76	0.38
Managing Momentum	5.00	0.70	0.13	0.52	0.89	0.31	1.77	0.34
Rules and routines	3.00	0.70	0.13	0.47	0.89	-0.27	2.20	0.46

Table 8

Categories of Disruptions (Lohmann, 2015)

Verbal.Disruption	Agitation	Lack.of.eagerness.to.learn
chatting with neighbor	drumming hands	putting head on table
whispering with neighbor	clicking pen	looking at phone
heckling	snipping with fingers	drawing on a sheet of paper

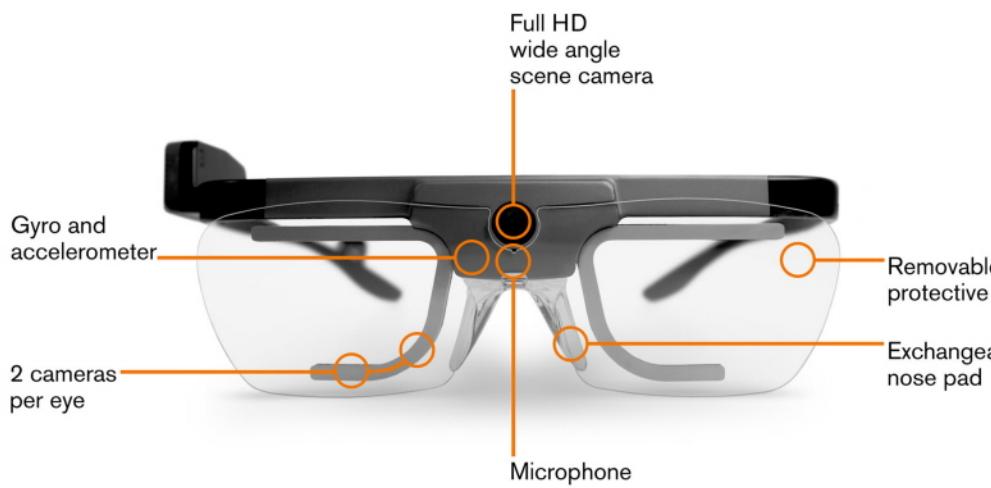


Figure 1. Tobii Pro Glasses 2; Source: <https://www.tobiipro.com/product-listing/tobii-pro-glasses-2/>

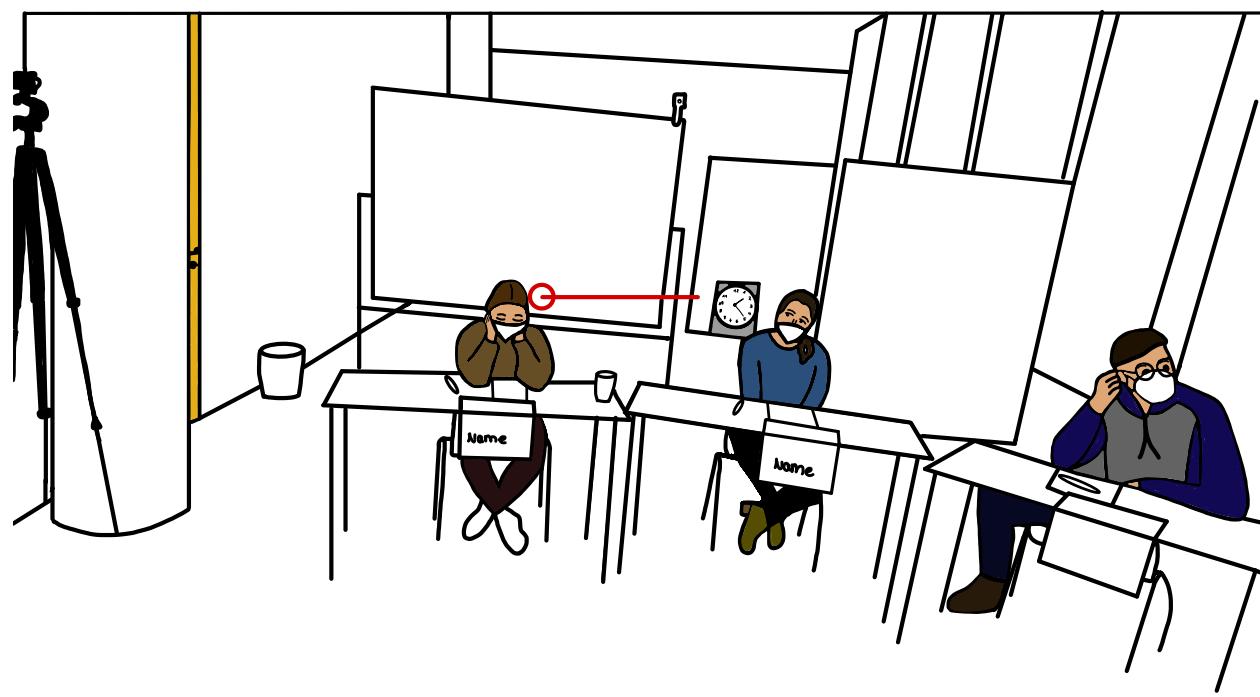


Figure 2. Teacher's Gaze Point



Figure 3. Subject and experimenter during the Stimulated Recall Interview