

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, 2006; Kounin, 2006). Accordingly, teachers must be able to quickly
17 notice and appropriately react to significant events in the classroom. This ability is
18 referred to as classroom professional vision (Goodwin, 1994; Sherin, 2007). The process of
19 professional vision can be divided into two main aspects: focusing on relevant situations for
20 learning and teaching (“noticing”) and applying knowledge to draw appropriate conclusions
21 in these situations (“knowledge-based reasoning”; Seidel & Stürmer, 2014). Therefore, the
22 early visual perception of classroom disruptions is a key component to effectively maximize
23 students’ learning time and minimize classroom interruptions. According to Kounin (2006),
24 these important classroom management strategies are called “withitness” and “overlapping”
25 and can be summarized under the concept of monitoring (Gold & Holodynski, 2017).
26 Learning to develop such classroom management skills is a demanding and complex task
27 for student teachers (Wolff et al., 2017). Research on teacher expertise showed that expert
28 and novice teachers differ in their ability to perceive classroom events, “[...] whereas only
29 a few studies have focused on the basal process of noticing, i.e. the recognition of possible
30 disturbing situations” (Grub et al., 2020, p.75). Mobile eye-tracking data can fill this
31 research gap by providing new insights in how expertise differences in teacher’s professional
32 vision manifest in teacher-student interactions (Lachner et al., 2016; Wolff et al., 2016).

33

Theoretical background

34 Professional competence in teaching profession

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

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

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

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

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

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

86 even offset the influence of an ineffective school setting (Marzano et al., 2003). This effect
87 places teachers at the center of ensuring effective teaching and learning and underscores the
88 importance of professional competence in this career field.

89 **Classroom Management**

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

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

109 Classroom management significantly influences the ability to perceive and
110 subsequently respond to classroom events and thus organize for disruption-free instruction

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

121 An initial explanatory approach to classroom management was provided by Doyle
122 (1990, p.355) with his definition of teaching, which he described as a "cognitive activity
123 based on knowledge of the probable course of events in the classroom and the ways in
124 which actions affect situations." From this, it is clear that in order to teach successfully, a
125 teacher needs knowledge of how to process as well as represent the full range of classroom
126 events. In addition, it is necessary to be able to respond or interact proactively and
127 effectively to potential events at best.

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

138 experience, which is why they constitute an essential component of teacher expertise (Wolff
139 et al., 2015).

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

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

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

162 Classroom management itself is made up of various components which, according to
163 Helmke (2014), can be divided into four categories: Rules and Procedures, presence, Time

- ¹⁶⁴ Management, and Establishing Desirable and Reducing Undesirable Student Behavior.
¹⁶⁵ Rules are considered the non-plus ultra of proactive management in the classroom, to
¹⁶⁶ whose agreements students and teachers alike must adhere (Helmke & Helmke, 2015).

¹⁶⁷ The importance of rules in the classroom has been known and researched for some
¹⁶⁸ time. Doyle (1989) relied as early as the late 1980s on studies showing that effective
¹⁶⁹ classroom management is due to timely implemented rules (cf. Emmer, Evertson &
¹⁷⁰ Anderson, 1980; Emmer, Sanford, Clements & Martin, 1982; cited in Doyle, 1989, p. 15).
¹⁷¹ Although he also stresses that establishing a system of rules is a difficult task. Classroom
¹⁷² rules are situation-dependent, as certain agreements only apply to certain phases of the
¹⁷³ lesson (for example, speaking quietly during group work; Doyle, 1989). Furthermore, order
¹⁷⁴ is always established jointly, in cooperation with the learners (Doyle, 1989).

¹⁷⁵ As a result, students significantly influence these rules by their willingness to follow
¹⁷⁶ them. Procedures are another tool to maintain the flow of the lesson, as recurring routines
¹⁷⁷ provide learners with security and clarity (Helmke & Helmke, 2015). This inevitably
¹⁷⁸ reduces the teacher's workload, eliminates additional explanations are no longer necessary,
¹⁷⁹ and it is clear to the learners at all times what needs to be done. needs to be done. In
¹⁸⁰ addition, teachers can present instructions for corresponding routines through non-verbal
¹⁸¹ actions to save time (Helmke, 2014). For the reasons just explained, it is clear
¹⁸² why Hattie (2013, p. 122) gives this component of classroom classroom management an
¹⁸³ effect size of $d = 0.76$.

¹⁸⁴ No other aspect of classroom management is as effective for classroom interruptions
¹⁸⁵ as the teacher's omnipresence (Hattie, 2013, p. 122); the effect size is $d = 1.42$. (Helmke &
¹⁸⁶ Helmke, 2015, p. 6). The behavioral pattern of presence is not a new concept. It was
¹⁸⁷ already recorded by Kounin in 1970 under the term *withitness*. Later research confirmed
¹⁸⁸ and validated his assumptions and definition of the term (cf. Helmke, 2014, p.10). In order
¹⁸⁹ to be present, the teacher must be able to control the classroom adequately and

superficially effectively at all times. This means she must be able to recognize relevant cues or situations that require possible action (Wolff et al., 2015). Presence can consequently be illustrated very vividly by means of the image of “eyes in the back of the head” (Kounin, 2006). In addition to presence, Kounin (2006) uses the term overlap. It is a kind of multitasking ability to be able to cope with parallel processes and situations in the classroom. The two dimensions of overlap and presence can also be summarized together under the term monitoring (Gold & Holodynski, 2017). When talking about ongoing monitoring in the classroom, this process includes teachers’ awareness of concurrent events as well as their demonstration of this to learners (Grub et al., 2020.). Well-developed monitoring behaviors enable teachers to detect disruptions early and increase overall awareness. and increases general awareness of relevant events in the classroom (ibid.). From this point of view, it becomes clear why awareness is a core component of teacher competence.

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

In order to enforce desired student behavior or to reduce contradictory behavior, teachers can use various discipline-related interventions and also make use of negative sanctions and positive reinforcement (Helmke & Helmke, 2015). The authors emphasize that a healthy balance must be found when using these strategies. Over the years, this constructive approach has fallen into disrepute due to its behaviorist character. However, elements such as reinforcement, extinction or punishment are not negligible laws of learning

²¹⁷ psychology whose effectiveness has been proven by Hattie (2013), among others. These
²¹⁸ behavioral psychological measures have an effect size of $d = 0.76$ (cf. Hattie, 2013, p.215)
²¹⁹ and are therefore an effective means of reducing disruptions (Helmke & Helmke, 2015).

²²⁰ Researchers in pedagogy also point out, however, that it is not only pure knowledge
²²¹ about classroom management that determines its effectiveness and influences it (Helmke &
²²² Helmke, 2015). The teacher personality, which includes characteristics such as subjective
²²³ tolerance margins, i.e. at what point does one perceive a certain behavior as disruptive,
²²⁴ also plays a role. This includes further aspects such as emotion control, empathy or the
²²⁵ credibility of a teaching person that arises through appearance. These characteristics,
²²⁶ which occur in many ways, have a direct impact on the quality of teaching and classroom
²²⁷ management, as they influence the planning behavior, perception and reaction of teachers.

²²⁸ Professional vision

²²⁹ Competence assessment plays a central role in the teaching profession and its
²³⁰ training. It follows the approach of concentrating on action-relevant as well as
²³¹ lesson-related content in order to train corresponding competences for teaching practice
²³² (KMK, 2004). In order for teachers in heterogeneous classrooms to be able to filter relevant
²³³ teaching aspects, they need knowledge about the control of their attention processes
²³⁴ (Barth, 2017). Science summarizes this competence under the so-called professional vision,
²³⁵ which is now a widely empirically studied research topic (Jahn, Stürmer, Seidel & Prenzel
²³⁶ 2014). The concept of professional vision can originally be traced back to the
²³⁷ anthropologist Charles Goodwin and became known under the term Professional Vision
²³⁸ (Sherin, 2001). It provided the answer to Goodwin's question regarding the development of
²³⁹ a professional vision, in the context of the profession, compared to the vision of everyday
²⁴⁰ situations. Goodwin uses the comparison of an archaeologist and a layperson to
²⁴¹ figuratively describe the differences. In doing so, he investigated which ability it is that
²⁴² makes the archaeologist see and interpret a landscape differently than the layperson

243 (cf. Sherin, 2001, p. 75). What looks to the former merely like sand, rocks and stones, is to
244 the trained eye of the archaeologist a base of a column or architectural fragments.

245 Goodwin (1994, p.606) defines professional vision as “socially organized ways of seeing and
246 understanding events that are answerable to the distinctive interests of a particular social
247 group” (cited in Sherin, 2001, p. 75). It is assumed that one can see certain phenomena or
248 occurring events in a certain way due to one’s professional affiliation (Sherin, 2001).

249 Particularities are more easily recognized because the gaze becomes more trained over time
250 and thus profession-relevant features appear more distinctive. In the previous example of
251 the archaeologist, this would mean that he noticed variations in color, texture and sand
252 and saw a cluster of stones as a possible element of a larger structure (cf. Sherin, 2001,
253 p.75). This distinguishes experts significantly in their visions from non-experts in the field,
254 as they had learned to filter their environment for crucial features.

255 **Teachers’ professional vision.** Based on the assumption that the professional
256 vision can develop over time, Sherin (2001) postulates that teachers can also acquire a
257 professional vision. As teachers develop from novices to experts in their field, they
258 accumulate expertise in a variety of areas (cf. Sherin, 2001, p. 76). This can be seen,
259 among other things, in flexible reactions to unexpected events or the smooth
260 implementation of routines in the classroom (Berliner, 1994; Leinhardt & Greeno, 1986).
261 This type of action can be described as a skill that helps teachers make decisions about
262 what is relevant and what is not within the complex classroom setting (van Es & Sherin,
263 2002). Such decisions are based on professional knowledge that guarantees effective
264 teaching and learning in actual classroom situations (Jahn et al., 2014). Professional
265 awareness is therefore referred to as an indicator of how knowledge is represented in
266 teachers (Stürmer et al., 2017). It can also be seen as a bridge or link between knowledge
267 and action, or professional competence and behavior (Grub et al., 2020). It is important to
268 understand that professional vision is acquired and not inherent. Rather, it is subject to a
269 process of acquisition that is constantly progressing as teachers are exposed to events and

270 learn to interpret them (Wolff et al., 2016). Fundamentally, then, prospective teachers
271 must select what to focus their attention on in the classroom and how to understand these
272 events (Stürmer et al., 2017). Finally, teaching and learning take place in a complex
273 environment. Simultaneous events and situations cannot be given the same attention.
274 “Selective perception is necessary in order to distinguish important from unimportant
275 features and to focus attention on relevant aspects” (Barth, 2017, p.19). Furthermore,
276 teachers benefit from from a trained perception of problems, as this is a significant factor in
277 being able to preventively counteract preventive measures (Wettstein, 2013).

278 Basically, the complex process of professional vision is divided into three
279 sub-processes (Grub et al., 2020). First of all, the identification, the recognition, of an
280 event takes place. Then connections are made, which are finally evaluated by the teacher
281 using his or her professional knowledge. In research, these processes are often combined
282 and one speaks of noticing to describe recognition (Grub et al., 2020). Knowledge-based
283 reasoning includes the last two processes of applying knowledge to make appropriate
284 decisions (*ibid.*). During this process, it becomes clear that the recognition of relevant
285 situations, such as disruptions, precedes all others. This process is therefore considered
286 necessary, without which professional vision cannot take place (Grub et al., 2020). As
287 described in the chapter on competence, perceptions is also of particular importance in the
288 competence model according to Blömeke et al. (2015). This is because the process of
289 perception forms an essential basis for situation-specific skills of teachers (Grub et al.,
290 2020). Inferentially, recognition is inevitable for the perception of classroom events and
291 therefore highly relevant for effective classroom management (*ibid.*).

292 Even though the recognition of relevant situations in the classroom is the starting
293 point for all further subsequent actions, the processes of knowledge-based reasoning within
294 professional vision are not to be neglected. It is also said that this process is an indicator of
295 the quality of how knowledge is applied to a given situation (Grub et al., 2020), as teachers
296 need to make knowledge-based inferences. Qualitative research on knowledge-based

297 reasoning concluded that it can be subdivided into three further sub-processes. Sherin
298 (2007; cited in Barth, 2017, p. 20) made a first subdivision, which consists of the processes
299 of describing, evaluating and interpreting. When a teacher describes what she notices, she
300 concentrates on observable characteristics, i.e. important aspects of teaching (*ibid.*). When
301 evaluating, statements are made about the quality of the teaching actions, and when
302 interpreting, conclusions are drawn about possible mechanisms of the event (cf. Barth,
303 2017, p. 19). Seidel et al. (2010), on the other hand, focus on German-speaking countries in
304 their research on professional vision and modified Sherin's original concept. They also
305 divide knowledge-based-reasoning into three levels, but these differ: describing, explaining
306 and predicting (Barth, 2017). More precisely, these processes encompass the ability to
307 differentially describe instructional components that have an impact on learning; to make
308 theoretically sound explanations about instructional situations; and to predict the effects of
309 these events in relation to the teaching-learning process (cf. Seidel et al., 2010, p. 297).
310 This scientific work examines only the processes of noticing, which is why these two
311 approaches have been presented for the sake of completeness. However, it is not useful to
312 specify which classification of knowledge-based reasoning is referred to in arguments. is
313 referred to in argumentation.

314 Against this scientific background, it can be deduced that the two processes of
315 noticing and knowledge-based-reasoning in professional vision cannot be considered
316 separately from each other, but that they have a reciprocal effect (Sherin, 2007; cited in
317 Barth, 2017, p. 19). What is considered and perceived as relevant in a teacher's selective
318 vision (noticing) functions as the basis of the subsequent processes of describing,
319 evaluating, and interpreting (knowledge-based-reasoning). Conversely, by engaging with
320 these processes, the teacher's perception is sensitized. This leads to the fact that in the
321 future certain characteristics of this situation can be recognized rather.

322 Measuring professional vision: eye-tracking technology in research

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

346 Eye-tracking technology can be roughly divided into two types of devices: Stationary
347 and mobile systems (Blake, 2013). In research so far, mainly eye-tracking devices have

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

359 Eye-tracking research is based on several fundamental assumptions. The connection
360 between mental processing and eye movements has already been carefully investigated and
361 provides viable empirical evidence to date (cf. Blake, 2018, p.380). In this regard,
362 Hoffmann and Subramaniam (1995) obtained the finding that when the eyes are focused on
363 an object, this is associated with an attentional focus of the person. Based on these
364 findings, ET technology is based on the following two assumptions: The immediacy
365 assumption and the eye-mind assumption (cf. Rakoczi, 2012, p.87). The former defines
366 that cognitive evaluation processes begin immediately upon fixation on an object
367 (Cullipher, 2018). If a new stimulus attracts attention and is fixated, the observer starts
368 processing this new information (*ibid.*). The eye-mind hypothesis assumes that whatever is
369 fixated by the eye, is subsequently processed by the mind (*ibid.*). It thereby expansively
370 defines, that objects are visually focused only until they are cognitively evaluated or until
371 there is have been cognitively evaluated or there is still an interest (cf. Rakoczi, 2012,
372 p.87). However, a situational context must exist for the assumptions just explained to be
373 are valid (*ibid.*). Through these functions, it becomes clear that the intrinsic processes of
374 recognition in professional vision can be measured by eye-tracking technology.

375 Despite the externalization of internalized processes and the resulting multitude of
376 data potentially to be collected, ET technology cannot provide answers about the exact
377 thoughts during stimulus processing. ET can identify where and when people are looking,
378 but it does not answer why. For this, it is helpful to triangulate the data and apply
379 alternative elicitation (Cullipher, 2018). One proven method is to conduct reflective
380 interviews, which can be conducted in the format of a think aloud protocol (Konrad, 2020).

381 This thinking aloud opens up the possibility of gaining insights into subjects' thinking,
382 feeling, and intentions (edb.). An elaborated form of the Think Aloud Protocol is a
383 Stimulated-Recall Interview (SRI). This is a qualitative scientific method to obtain data on
384 preactional thought processes (cf. Messmer, 2015, p.4). The pre-recorded lesson sequences
385 are played back and subjects have the opportunity to express themselves about their
386 "thoughts, beliefs, norms, and decisions" (Messmer, 2015, p.4) during the interview
387 process. Thus, the visual data of the ET can be matched with cognitive processes and
388 processes of knowledge-based-reasoning can be better understood. Individual-dependent
389 thinking is now no longer pure speculation and can be made objectively observable in a
390 way (Messmer, 2015). Subjects also have the opportunity to explain their actions and
391 become aware of the basis for their decisions (ibid.).

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

399 In MET, collected data sets are usually output according to different parameters:
400 Fixations and Saccades (Gegenfurtner et al., 2018). The latter describes gaze jumps, i.e.,
401 movements of both eyes, which are useful to capture new fixations (Eye-Tracking

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

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

422 **Expertise differences in professional vision.** Since it is assumed that
423 professional vision is a preferably knowledge-based process, differences between experienced
424 and inexperienced teachers have already been identified. These differences were in favor of
425 experts, since a connection between teaching experience and the associated increase in
426 knowledge acquisition is assumed (Grub et al., 2020). Student teachers do not have
427 well-structured experiential knowledge due to their almost non-existent practice in
428 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, 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, Baumert and Kunter (2006) 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, 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 such as classroom disruptions. In the worst case, they force a standstill in the flow of instruction, which is naturally to be avoided. Lohmann (2015, p.13) describes classroom

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

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

478 In the last decade, visual input through recorded teaching situations has been
479 preferred (*ibid.*). Subjects are thus shown an outside perspective from different points of
480 view. This can happen either as self or external video and allows micro-analysis or
481 interaction sequences through the characteristic of repeated playing of a video (*ibid.*). This
482 characteristic makes it possible to analyze corresponding scenes more intensively and in

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

502 A successful method to reduce this hindering factor of self-drawn cognition and still
503 use self videos for teacher education is to use recordings that show the teacher's perspective
504 and furthermore make it known where the teacher was looking (cf. Cortina et al., 2018). In
505 MET's video material, teachers can never see their teaching through their own eyes. This
506 makes it possible to focus more specifically on essential aspects, such as classroom
507 disruptions and interactions with learners. Although it has not yet been clearly
508 demonstrated that significantly less self-referential cognition was elicited, self-videos
509 counteract over-focusing on the acting person (cf. Cortina et al., 2018). Therefore, MET

510 represents a target-rich method for capturing teachers' attentional processes in the
511 classroom. Previous studies with large data sets on teachers' visual expertise have
512 predominantly used stationary ET (Wolff et al., 2016). In MET studies, the samples turn
513 out to be much smaller; here, the largest study to be mentioned is with 12 experienced and
514 12 inexperienced teachers by Cortina et al. (2015).

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

537 such as self-efficacy beliefs (ibid. (ibid.). With decreasing self-efficacy, an increasing sense
538 of disturbance was found (ibid.).

539 The teacher-student relationship aspect, which includes recognizing student needs,
540 must also be considered. In order to be able to adequately counteract classroom
541 disruptions and thus automatically reduce the sense of disruption and the experience of
542 stress, it is important to address the concerns of the learners with empathy (Wickeren,
543 2019). Regarding gender as a factor, it should be noted that previous research has found
544 higher levels of disruptiveness in discipline problems among female teachers
545 (Kovess-Masféty, Rios-Seidel & Sevilla-Dedieu, 2007; cited in Wickeren, 2019, p. 33). In
546 summary can thus be summarized that the extent and perception of disturbances are based
547 on individual cognitive evaluation processes, the situational context, attentional focus,
548 didactic staging, and, indeed, professional experience (cf. Wickeren, 2019).

549 **Research Questions and hypothesis**

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

558 Based on the existing literature, we expected expert teachers to outperform novices
559 by (H1) showing more fixations on relevant areas with (H2) shorter fixation duration and
560 (H3) perceiving classroom disruptions faster (cf. Van den Bogert, Bruggen, Kostons, and
561 Jochems (2014)), (H4) feeling less troubled by disruptions and (H5) more confident in

562 dealing with disruptions (Barth, 2017), and (H6) having higher scores in strategic
563 knowledge of classroom management (Gold et al., 2016).

564 **Materials and Methods**

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

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

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

580 **Participants**

581 The sample consisted of $N = 45$ participants with $n = 19$ expert teachers and $n = 26$
582 novice teachers.

583 The inclusion criterion for experts was that they have successfully completed teacher
584 training and are actively employed in the teaching profession. According to Palmer,
585 Stough, Burdenski, and Gonzales (2005), we selected teachers as experts who had at least

586 three years of professional experience and ideally had worked in another teaching position,
587 such as subject advisor or trainer for trainee teachers, in addition to their teaching
588 profession in school. Novices were student teachers who had successfully completed their
589 first internship in a school and gained one to four hours of teaching experience.

590 The expert teachers (11 women; 57.89%) had a mean age of 41.70 years ($SD = 10.40$;
591 range: 27-59) and an average teaching experience of 13.80 years ($SD = 12.30$; range: 3-37).
592 15.79% of the experts were primary school teachers and 84.21% were secondary school
593 teachers. 47% of the experienced teachers were also engaged in an secondary teaching
594 activity, such as lecturers at the university, main training supervisors for trainee teachers
595 and subject advisers.

596 The novice teachers (17 women; 65.38%) had a mean age of 23.40 years ($SD = 1.70$;
597 range: 20-27) with an average teaching experience of 0 years. On average, the student
598 teachers were in their 7.20 semester ($SD = 2.60$; range: 3-11). Furthermore, they had an
599 average teaching experience of 10.90 teaching units à 45min ($SD = 8.10$; range: 0-36)
600 through the internships during their studies. 19.23% of the novices were studying to
601 become primary school teachers, 65.38% to become secondary school teachers and 15.38%
602 to become special education teachers. 88.46% of the student teachers were also engaged in
603 an extracurricular teaching activity, such as tutoring or homework supervision.

604 The subjects were primarily recruited through personal contacts, social media
605 (Facebook), e-mail distribution lists and advertising in lectures at the University Leipzig.
606 All study procedures were carried out in accordance with the ethical standards of the
607 University's Institutional Review Board. The authors received a positive vote on the study
608 procedures from the Ethics Committee Board of Leipzig University. All participants were
609 informed in detail about the aim and intention of the study prior to testing. Participation
610 in the study was voluntary and only took place after written consent has been given.

611 **Design**

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

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

629 **Materials**

630 **Eye-Tracking Apparatus.** To record eye-tracking data, teachers wore a binocular
631 Tobii Pro Glasses 2 eye-tracker during the micro-teaching-unit. The system consisted of a
632 wearable head unit and a recording unit. As shown in Figure 1, the head unit was a
633 measuring device with different sensors. A high-definition scene camera captured a full HD
634 video of the teacher's field of vision. An integrated microphone recorded the surrounding
635 sounds. Infrared light illuminators supported the eye tracking sensors which recorded the

636 eye orientation to capture the teacher's gaze point as shown in Figure 2. The videos were
637 recorded with a sampling rate of 50 Hz in a video resolution with 1920 x 1080 at 25 frames
638 per second. The scene camera had a field of view of 90 deg. in 16:9 format (82 deg.
639 horizontal and 52 deg. vertical) and a frame dimension of 179 x 159 x 57 mm (width x
640 depth x height).

641 **Video data.**

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

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

656 Video analysis was chosen as the research method for evaluating the teacher's
657 reaction to classroom disruptions because it offers numerous advantages. In contrast to
658 transcripts or audio-only recordings, a video also shows nonverbal forms of communication,
659 which is emphasized here, as well as the entire complex lesson excerpt in which many
660 things happen simultaneously (cf. Krammer & Reusser, 2012, p. 47f). Video data can also
661 be viewed and analyzed over and over again from different perspectives (Petko et al., 2003,

662 p. 265). The possible problem of selectivity (cf. Petko et al., 2003, p. 271), i.e., the fact
663 that only a section of the events can be shown due to the position and perspective of the
664 camera, was countered by recording the lessons from a total of four cameras. In this way, a
665 case of doubt can be analyzed from several angles. Another challenge can be the so-called
666 ‘camera effect’, i.e. that the filmed persons feel disturbed and act differently than usual and
667 less naturally. However, Petko et al. (2003) conclude, despite filming with a camera person
668 at the time, that there is little reason to believe that teachers can easily change their
669 teaching style from one lesson to the next (p. 270) and that it should not be assumed that
670 the mere presence of a camera has significantly improved or worsened teaching’ (p. 270). In
671 the ‘ProVisioNET’ survey, four very small cameras were placed in the room and their
672 presence was barely noticeable. Accordingly, it is assumed that no major nuisance
673 emanates from them.

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

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

683 A theory-based coding scheme was developed as an observation tool for classroom
684 observation. The coding of the reactions was done in seven levels. Based on the literature
685 on effective approaches to classroom disruptions, nonverbal response was given Code 1 as
686 the best variant because it best accommodates the classroom management strategies of
687 presence and overlapping (cf. Kounin 2006; Gold & Holodynski 2015) by not interrupting
688 the flow of instruction. Code 2 or Code 3 was assigned for a brief verbal expression with or

689 without a preceding nonverbal response. A brief, timely, and purposeful rebuke of the
690 disruptive student is considered by Kounin (2009, p. 99), among others, to be an important
691 component of the strategy of presence and thus demonstrates good classroom management.
692 If the teacher reacts non-verbally and at the same time with a longer verbal statement,
693 code 4 was assigned, because the flow of the lesson was already noticeably delayed. There
694 were two possibilities, if there was no reaction: If the teacher deliberately ignored the
695 disruption, code 5 was assigned. If the teacher did not notice the disruption, code 6 was
696 assigned, because then no effective monitoring was realized (cf. Gold & Holodynksi, 2015).
697 If the reaction to the disruption was dysfunctional, i.e., if the teacher shows an
698 inappropriate reaction, such as explaining the class rules for a very long time in the case of
699 an insignificant disruption, code 7 was assigned. According to Gippert et al. (2019), there
700 is a relational error in such cases, as the intensity of the student response should be relative
701 to the disruptive behavior (p. 2).

702 In developing the coding system, questions initially aroused about how descriptions of
703 responses should be made and in what order (especially codes 3-4). Initially, it was
704 planned to assign code 3 for a nonverbal and simultaneously long verbal response and code
705 4 for a purely verbal response. However, during the initial trial coding of the videos, it
706 became apparent that it was difficult to reliably distinguish nonverbal reactions to a
707 classroom disruption from normal facial expressions and gestures of the teacher when
708 talking. Moreover, during a prolonged verbal utterance, nonverbal signs could virtually
709 always be detected. Furthermore, a category for short, purely verbal reactions was missing.

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

713 Test quality criteria

714 In the construction of the video analysis method, attention was paid to compliance

715 with the quality criteria of objectivity, reliability and validity.

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

728 With the help of scripted behavioral instructions for students in the instructional
729 unit, a high degree of standardization was achieved that is not present in the normal school
730 setting. The lessons were recorded with several video cameras from different angles and an
731 audio recorder. This means a big advantage compared to the subjective evaluation of
732 observers in real time. In addition, video recordings can be viewed multiple times, by
733 different people or from different perspectives. Thus, a high measurement accuracy
734 (reliability) is given (cf. Mayring, P., Gläser-Zikuda, M., & Ziegelbauer, 2005). All
735 measures taken contribute to increase the validity of the method. However, it cannot be
736 answered whether the coding scheme is completely valid, because a validation with the help
737 of another measuring instrument was not possible. Therefore, it could not be controlled
738 whether the collected construct is related to other similar constructs.

739 *Coding Time on Task.*

740 **Questionnaire data.**

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

748 XXX FRANZI

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

754 The questionnaire included nine questions for each scripted disruptive event. The
755 subjects did not have to complete the questionnaire themselves at any time, as this was
756 done by the experimenter.

757 For the present study, the following two questions from the interview were relevant:
758 1. "How disturbing did you perceive this event to be?" 2. "How confident did you feel in
759 dealing with this event?"

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

763 With these rating scales, the subjects indicated their personal subjective feeling
764 towards the individual lesson disruption. Rating scales usually consist of a number of
765 clearly arranged categories that can be presented visually in a variety of ways (Albers,
766 Klapper, Konradt, Walter, & Wolf, 2009, p. 67f). In the present questionnaire, the

767 numerical representation with eleven answer options (0-10) was used. There was no specific
768 designated neutral position. In this questionnaire, the middle category 5 was not
769 interpreted as a neutral position. In order to avoid misinterpretations of the scaling, the
770 questionnaire was described verbally by the experimenter and follow-up questions from the
771 test subjects were possible at any time. With regard to the quality criteria, it should be
772 made clear that complete objectivity in answering the two questions is given by the rating
773 scales. The question remained the same for all subjects and no additional comments were
774 made. Validity can also be assumed for the questionnaire, since the subjects could ask
775 follow-up questions and thus uncertainties were clarified. However, similar to the video
776 data, there was no perfect reliability with this survey instrument, because in a few
777 exceptional cases, for example, the experimenter forgot to ask the subject the relevant
778 questions.

779 *Situational Jugdement Test of Strategic Knowledge of Classroom*

780 **Management.** To assess the strategic knowledge of classroom management, subjects
781 answered a Situational Judgement Test (SJT, (Gold & Holodynski, 2015)). In the
782 questionnaire, participants were given hypothetical school scenarios with five to six
783 response strategies for which they had to evaluate the effectiveness of each strategy on a
784 6-point rating scale from 1 (A) to 6 (F) according to school grades (see example scenario
785 XXX REFERENZ EINFÜGEN). Depending on the grade given to each strategy, scores
786 were converted in an SPSS syntax resulting in scores ranging from 0 to 1 for each
787 participant.

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

791 *Monitoring* includes proactive strategies such as omnipresence. This means, the
792 teacher is always informed about the students' behavior, regulates it and reports it back to
793 them. The second strategy is overlapping. In this case, the teacher is able to control

794 several teaching processes at the same time. Monitoring means also reactive strategies of
795 supervision, which also includes the effective handling of classroom disruptions (cf. ibid.;
796 [Kounin (2006), p. 89ff].

797 By *managing momentum* (structuring), Gold and Holodynski (2015) understand the
798 skillful control of teaching processes so that they can run smoothly and without delays,
799 taking into account the students' level of understanding and attention (Kounin, 2006, p.
800 101ff). They also include the group focus specifically identified by Kounin (Kounin, 2006,
801 p. 117ff), i.e. mobilizing the whole class and demanding accountability (Gold &
802 Holodynski, 2015, p. 231).

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

806 The construction of the SJT included 14 scenarios that were based on transcribed
807 classroom videos from Germany. For the 14 scenarios, response strategies were derived
808 from evidence-based and theoretical principles by 17 experts on classroom management.
809 With the exception of scenario 5, which was subsequently excluded, the 17 experts agreed
810 on an appropriate level of content validity for all other scenarios (cf. ibid.). For the
811 validation of the strategies, pair comparisons were made between different strategies and
812 the frequencies were calculated in relation to the expert responses (cf. ibid., p. 238).
813 Scenarios 10 and 12 were excluded from further analysis because they did not reach the
814 minimum level of agreement (cf. ibid.).

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

820 validity of the SJT test procedure was assumed.

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

824 In the current study, objectivity of application was ensured by the
825 questionnaire-format, the written instructions and by a detailed test administration
826 manual of the study *ProVisioNET*. The objectivity of analysis was given by processing the
827 obtained questionnaire data in R (Version 4.1.3; R Core Team, 2021) and the R-packages
828 *ARTofR* (Version 0.4.1; Zhang, 2021), *cowplot* (Version 1.1.1; Wilke, 2020), *DescTools*
829 (Version 0.99.45; Andri et mult. al., 2022), *dplyr* (Version 1.0.8; Wickham, François, Henry,
830 & Müller, 2022), *forcats* (Version 0.5.1; Wickham, 2021), *formattable* (Version 0.2.1; Ren &
831 Russell, 2021), *ggridmosaic* (Version 0.3.3; Jeppson, Hofmann, & Cook, 2021), *ggplot2*
832 (Version 3.3.5; Wickham, 2016), *ggrepel* (Version 0.9.1; Slowikowski, 2021), *ggthemes*
833 (Version 4.2.4; Arnold, 2021), *gridExtra* (Version 2.3; Auguie, 2017), *haven* (Version 2.4.3;
834 Wickham & Miller, 2021), *kableExtra* (Version 1.3.4; Zhu, 2021), *knitr* (Version 1.38; Xie,
835 2015), *lme4* (Version 1.1.29; Bates, Mächler, Bolker, & Walker, 2015), *ltm* (Version 1.2.0;
836 Rizopoulos, 2006), *lubridate* (Version 1.8.0; Grolemund & Wickham, 2011), *MASS* (Version
837 7.3.55; Venables & Ripley, 2002), *Matrix* (Version 1.4.0; Bates & Maechler, 2021), *moments*
838 (Version 0.14; Komsta & Novomestky, 2015), *msm* (Version 1.6.9; Jackson, 2011), *needs*
839 (Version 0.0.3; Katz, 2016), *papaja* (Version 0.1.0.9999; Aust & Barth, 2020), *polycor*
840 (Version 0.8.1; Fox, 2022), *purrrr* (Version 0.3.4; Henry & Wickham, 2020), *readr* (Version
841 2.1.2; Wickham, Hester, & Bryan, 2021), *readxl* (Version 1.4.0; Wickham & Bryan, 2019),
842 *rgeos* (Version 0.5.9; Bivand & Rundel, 2021), *rlang* (Version 1.0.2; Henry & Wickham,
843 2022), *rnatural-earth* (Version 0.1.0; South, 2017a, 2017b), *rnatural-earthdata* (Version 0.1.0;
844 South, 2017b), *sf* (Version 1.0.7; E. Pebesma, 2018), *sjlabelled* (Version 1.2.0; Lüdecke,
845 2022), *sjPlot* (Version 2.8.10; Lüdecke, 2021), *sp* (Version 1.4.7; E. J. Pebesma & Bivand,
846 2005), *stringr* (Version 1.4.0; Wickham, 2019), *tibble* (Version 3.1.6; Müller & Wickham,

847 2021), *tidyverse* (Version 1.3.1; Wickham & Girlich, 2022), *tidyverse* (Version 1.3.1; Wickham
848 et al., 2019), *tinylabels* (Version 0.2.3; Barth, 2022), *viridis* (Version 0.6.2; Garnier et al.,
849 2021a, 2021b), *viridisLite* (Version 0.4.0; Garnier et al., 2021b), and *xtable* (Version 1.8.4;
850 Dahl, Scott, Roosen, Magnusson, & Swinton, 2019) and IBM SPSS 28.

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

856 XXX FRANZI

857 Procedure

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

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

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

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

871 After the subject arrived, a smart watch was put on to measure the heart rate during

872 the session and to get a pretest time at least 15min before the session started. In addition

873 to the experimenter and the subject, three student assistants from the working group

874 always took part in the data collection, as they represented the class.

875 After the welcome, the subject was again briefed about the study. It was explicitly

876 pointed out that the student assistants would act as the class and simulate typical class

877 events during the lesson. The subject was asked in advance to behave as naturally as

878 possible during the entire time. Next, the subjects' written informed consent was obtained

879 and contact details were collected in order to inform all persons participating in the study

880 if a covid infection should occur.

881 After the introduction, the eye-tracking glasses were adjusted for the subject by

882 inserting contact lenses if necessary and changing the nose pad. To start the eye tracking

883 glasses, the Tobii Glasses must be fitted onto the subject's head via an

884 one-point-calibration. In the calibration process the subject was asked to look at a

885 Calibration Card held in-front of the subject for a few seconds. The experimenter started

886 the recording from Tobii Glasses Controller Software running on a computer.

887 After starting the eye tracking recording, all other technical devices were also

888 switched on: The four cameras and the audio recorder were controlled via iPad using the

889 BLINK Hup app and could be started simultaneously by synchronization. The ZED

890 camera was started manually on another laptop.

891 Before the 15-minute lesson, there was a short 10-minute warm-up phase. The phase

892 was divided into two parts and served on the one hand to get the subjects used to the

893 eye-tracking glasses and on the other hand to get used to their class. In the first phase of

894 the warm-up, the game "name juggling" was played using two balls. In the game, the

895 subject and the three actors threw two balls at each other and, depending on the color of

896 the ball, called either their own name or that of the target person. After the name

897 juggling, the subjects were supposed to start a conversation. For this, the subject thought
898 of a question for each student and was also asked a question of each student. The content
899 could be anything that interested the participants.

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

910 For the micro-teaching lesson, student teachers and experienced teachers were asked
911 to prepare an introduction of 15 minutes which they taught in front of the fictitious class
912 consisting of three student assistants. The actors simulated the nine classroom events
913 during the lesson, derived by research literature. The order of the disruptions as well as the
914 students performing them were fully balanced using Latin Square. The disruptions were
915 only visible for the class on a screen.

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

920 After the lesson, the time was again noted from the smartwatch as well as the
921 subject's steps. The nine-point calibration was also performed again in the neighboring
922 room. This time, however, the subject was asked to wait outside the room until he or she

923 was called in, because four letters from A to D were placed in the study room. The subject
924 was asked to stand facing the board at a marked point in the room and, when given an
925 acoustic signal, to turn around and search the letters and read them aloud in order. This
926 served as a control condition for the speed of the subjects' perceptual ability as no
927 expertise is required for searching letters.

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

935 In the second part of the study, the experimenter conducted a Stimulated Recall
936 Interview (SRI) with the subject(see in Figure 3. The interview was recorded by the
937 ambient recorder and additionally with a software called OBS to record the speech and
938 screen. The recorded eye-tracking video was watched and commented on by the subject
939 while thinking aloud. The interview was structured by a guideline based on a fixed
940 sequence of questions, however the subjects were allowed to answer completely open to
941 these questions. The objective was to collect data that would provide insight into the
942 subjects' thoughts and reactions in relation to the scripted classroom disruptions. The
943 eye-tracking video was not completely re-watched, but stopped at the points where
944 disruptions occurred. After the disruption occurred, the subject was asked to describe
945 what had just happened. If the disruption was noticed, the subject was asked how
946 disruptive the event was at that moment on a scale from 0 to 10. In addition, the subject
947 was asked to reason the given number. In a next step, the experimenter re-played the
948 subject's reaction to the disruption and asked to describe and reason the reaction. The
949 experimenter then asked the subject to rate on a scale from 0 to 10 how confident he/she

950 felt in dealing with the disruption and again to reason the number (for a detailed
951 description of the scale question, see Rating Scales).

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

957 Data analysis

958 We investigated whether experts and novice teachers differed

959 All reported data analyses were conducted with the R (Version 4.1.3; R Core Team,
960 2021) and the R-packages *ARTofR* (Version 0.4.1; Zhang, 2021), *cowplot* (Version 1.1.1;
961 Wilke, 2020), *DescTools* (Version 0.99.45; Andri et mult. al., 2022), *dplyr* (Version 1.0.8;
962 Wickham et al., 2022), *forcats* (Version 0.5.1; Wickham, 2021), *formattable* (Version 0.2.1;
963 Ren & Russell, 2021), *ggridmosaic* (Version 0.3.3; Jeppson et al., 2021), *ggplot2* (Version
964 3.3.5; Wickham, 2016), *ggrepel* (Version 0.9.1; Slowikowski, 2021), *ggthemes* (Version 4.2.4;
965 Arnold, 2021), *gridExtra* (Version 2.3; Auguie, 2017), *haven* (Version 2.4.3; Wickham &
966 Miller, 2021), *kableExtra* (Version 1.3.4; Zhu, 2021), *knitr* (Version 1.38; Xie, 2015), *lme4*
967 (Version 1.1.29; Bates et al., 2015), *ltm* (Version 1.2.0; Rizopoulos, 2006), *lubridate*
968 (Version 1.8.0; Grolemund & Wickham, 2011), *MASS* (Version 7.3.55; Venables & Ripley,
969 2002), *Matrix* (Version 1.4.0; Bates & Maechler, 2021), *moments* (Version 0.14; Komsta &
970 Novomestky, 2015), *msm* (Version 1.6.9; Jackson, 2011), *needs* (Version 0.0.3; Katz, 2016),
971 *papaja* (Version 0.1.0.9999; Aust & Barth, 2020), *polycor* (Version 0.8.1; Fox, 2022), *purrr*
972 (Version 0.3.4; Henry & Wickham, 2020), *readr* (Version 2.1.2; Wickham et al., 2021),
973 *readxl* (Version 1.4.0; Wickham & Bryan, 2019), *rgeos* (Version 0.5.9; Bivand & Rundel,
974 2021), *rlang* (Version 1.0.2; Henry & Wickham, 2022), *rnatnarearth* (Version 0.1.0; South,

975 2017a, 2017b), *rnaturrearthdata* (Version 0.1.0; South, 2017b), *sf* (Version 1.0.7; E.
976 Pebesma, 2018), *sjlabelled* (Version 1.2.0; Lüdecke, 2022), *sjPlot* (Version 2.8.10; Lüdecke,
977 2021), *sp* (Version 1.4.7; E. J. Pebesma & Bivand, 2005), *stringr* (Version 1.4.0; Wickham,
978 2019), *tibble* (Version 3.1.6; Müller & Wickham, 2021), *tidyverse* (Version 1.3.1; Wickham et al., 2019), *tinylabels* (Version 0.2.3;
979 Girlich, 2022), *viridis* (Version 0.6.2; Garnier et al., 2021a, 2021b), *viridisLite* (Version 0.4.0;
980 Barth, 2022), *xtable* (Version 1.8.4; Dahl et al., 2019) and IBM SPSS 28.

982

Results

983

Discussion

References

- 984
985 Albers, S., Klapper, D., Konradt, U., Walter, A., & Wolf, J. (2009). *Methodik der*
986 *empirischen forschung* (Vol. 3). Springer.
- 987 Andri et mult. al., S. (2022). *DescTools: Tools for descriptive statistics*. Retrieved
988 from <https://cran.r-project.org/package=DescTools>
- 989 Arnold, J. B. (2021). *Ggthemes: Extra themes, scales and geoms for 'ggplot2'*.
990 Retrieved from <https://CRAN.R-project.org/package=ggthemes>
- 991 Auguie, B. (2017). *gridExtra: Miscellaneous functions for "grid" graphics*. Retrieved
992 from <https://CRAN.R-project.org/package=gridExtra>
- 993 Aust, F., & Barth, M. (2020). *papaja: Prepare reproducible APA journal articles*
994 *with R Markdown*. Retrieved from <https://github.com/crsh/papaja>
- 995 Barth, M. (2022). *tinylabels: Lightweight variable labels*. Retrieved from
996 <https://cran.r-project.org/package=tinylabels>
- 997 Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects
998 models using lme4. *Journal of Statistical Software*, 67(1), 1–48.
999 <https://doi.org/10.18637/jss.v067.i01>
- 1000 Bates, D., & Maechler, M. (2021). *Matrix: Sparse and dense matrix classes and*
1001 *methods*. Retrieved from <https://CRAN.R-project.org/package=Matrix>
- 1002 Bivand, R., & Rundel, C. (2021). *Rgeos: Interface to geometry engine - open source*
1003 *('GEOS')*. Retrieved from <https://CRAN.R-project.org/package=rgeos>
- 1004 Dahl, D. B., Scott, D., Roosen, C., Magnusson, A., & Swinton, J. (2019). *Xtable:*
1005 *Export tables to LaTeX or HTML*. Retrieved from
1006 <https://CRAN.R-project.org/package=xtable>
- 1007 Fox, J. (2022). *Polycor: Polychoric and polyserial correlations*. Retrieved from
1008 <https://CRAN.R-project.org/package=polycor>
- 1009 Garnier, Simon, Ross, Noam, Rudis, Robert, ... Cédric. (2021a). *viridis -*
1010 *colorblind-friendly color maps for r*. <https://doi.org/10.5281/zenodo.4679424>

- 1011 Garnier, Simon, Ross, Noam, Rudis, Robert, ... Cédric. (2021b). *viridis -*
1012 *colorblind-friendly color maps for r.* <https://doi.org/10.5281/zenodo.4679424>
- 1013 Gegenfurtner, A., Eichinger, A., Latzel, R., Dietrich, M. P., Barkowsky, M., Glufke,
1014 A., ... Stern, W. (2018). *Mobiles eye-tracking in den angewandten*
1015 *wissenschaften.*
- 1016 Gold, B., & Holodynki, M. (2015). Development and construct validation of a
1017 situational judgment test of strategic knowledge of classroom management in
1018 elementary schools. *Educational Assessment, 20*(3), 226–248.
- 1019 Golemund, G., & Wickham, H. (2011). Dates and times made easy with lubridate.
1020 *Journal of Statistical Software, 40*(3), 1–25. Retrieved from
1021 <https://www.jstatsoft.org/v40/i03/>
- 1022 Helmke, A., Helmke, T., Lenske, G., Pham, G., Praetorius, A.-K., Schrader, F.-W.,
1023 & AdeThuow, M. (2014). Unterrichtsdiagnostik mit EMU. *Aus- Und Fortbildung*
1024 *Der Lehrkräfte in Hinblick Auf Verbesserung Der Diagnosefähigkeit, Umgang*
1025 *Mit Heterogenität Und Individuelle Förderung*, 149–163.
- 1026 Henry, L., & Wickham, H. (2020). *Purrr: Functional programming tools.* Retrieved
1027 from <https://CRAN.R-project.org/package=purrr>
- 1028 Henry, L., & Wickham, H. (2022). *Rlang: Functions for base types and core r and*
1029 *'tidyverse' features.* Retrieved from <https://CRAN.R-project.org/package=rlang>
- 1030 Jackson, C. H. (2011). Multi-state models for panel data: The msm package for R.
1031 *Journal of Statistical Software, 38*(8), 1–29. Retrieved from
1032 <https://www.jstatsoft.org/v38/i08/>
- 1033 Jeppson, H., Hofmann, H., & Cook, D. (2021). *Ggmosaic: Mosaic plots in the*
1034 *'ggplot2' framework.* Retrieved from
1035 <https://CRAN.R-project.org/package=ggmosaic>
- 1036 Katz, J. (2016). *Needs: Attaches and installs packages.* Retrieved from
1037 <https://CRAN.R-project.org/package=needs>

- 1038 Komsta, L., & Novomestky, F. (2015). *Moments: Moments, cumulants, skewness,*
1039 *kurtosis and related tests*. Retrieved from
1040 <https://CRAN.R-project.org/package=moments>
- 1041 Kounin, J. S. (2006). *Techniken der klassenführung*. Waxmann Verlag.
- 1042 Lüdecke, D. (2021). *sjPlot: Data visualization for statistics in social science*.
1043 Retrieved from <https://CRAN.R-project.org/package=sjPlot>
- 1044 Lüdecke, D. (2022). *Sjlabelled: Labelled data utility functions (version 1.2.0)*.
1045 <https://doi.org/10.5281/zenodo.1249215>
- 1046 Müller, K., & Wickham, H. (2021). *Tibble: Simple data frames*. Retrieved from
1047 <https://CRAN.R-project.org/package=tibble>
- 1048 Palmer, D. J., Stough, L. M., Burdenski, T. K., Jr, & Gonzales, M. (2005).
1049 Identifying teacher expertise: An examination of researchers' decision making.
1050 *Educational Psychologist*, 40(1), 13–25.
- 1051 Pebesma, E. (2018). Simple Features for R: Standardized Support for Spatial Vector
1052 Data. *The R Journal*, 10(1), 439–446. <https://doi.org/10.32614/RJ-2018-009>
- 1053 Pebesma, E. J., & Bivand, R. S. (2005). Classes and methods for spatial data in R.
1054 *R News*, 5(2), 9–13. Retrieved from <https://CRAN.R-project.org/doc/Rnews/>
- 1055 R Core Team. (2021). *R: A language and environment for statistical computing*.
1056 Vienna, Austria: R Foundation for Statistical Computing. Retrieved from
1057 <https://www.R-project.org/>
- 1058 Ren, K., & Russell, K. (2021). *Formattable: Create 'formattable' data structures*.
1059 Retrieved from <https://CRAN.R-project.org/package=formattable>
- 1060 Rizopoulos, D. (2006). Ltm: An r package for latent variable modelling and item
1061 response theory analyses. *Journal of Statistical Software*, 17(5), 1–25. Retrieved
1062 from <https://doi.org/10.18637/jss.v017.i05>
- 1063 Slowikowski, K. (2021). *Ggrepel: Automatically position non-overlapping text labels*
1064 *with 'ggplot2'*. Retrieved from <https://CRAN.R-project.org/package=ggrepel>

- 1065 South, A. (2017a). *Rnaturalearth: World map data from natural earth*. Retrieved
1066 from <https://CRAN.R-project.org/package=rnaturalearth>
- 1067 South, A. (2017b). *Rnaturalearthdata: World vector map data from natural earth
1068 used in 'rnaturalearth'*. Retrieved from
1069 <https://CRAN.R-project.org/package=rnaturalearthdata>
- 1070 Van den Bogert, N., Bruggen, J. van, Kostons, D., & Jochems, W. (2014). First
1071 steps into understanding teachers' visual perception of classroom events.
1072 *Teaching and Teacher Education*, 37, 208–216.
- 1073 Venables, W. N., & Ripley, B. D. (2002). *Modern applied statistics with s* (Fourth).
1074 New York: Springer. Retrieved from <https://www.stats.ox.ac.uk/pub/MASS4/>
- 1075 Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Springer-Verlag
1076 New York. Retrieved from <https://ggplot2.tidyverse.org>
- 1077 Wickham, H. (2019). *Stringr: Simple, consistent wrappers for common string
1078 operations*. Retrieved from <https://CRAN.R-project.org/package=stringr>
- 1079 Wickham, H. (2021). *Forcats: Tools for working with categorical variables (factors)*.
1080 Retrieved from <https://CRAN.R-project.org/package=forcats>
- 1081 Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., ...
1082 Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*,
1083 4(43), 1686. <https://doi.org/10.21105/joss.01686>
- 1084 Wickham, H., & Bryan, J. (2019). *Readxl: Read excel files*. Retrieved from
1085 <https://CRAN.R-project.org/package=readxl>
- 1086 Wickham, H., François, R., Henry, L., & Müller, K. (2022). *Dplyr: A grammar of
1087 data manipulation*. Retrieved from <https://CRAN.R-project.org/package=dplyr>
- 1088 Wickham, H., & Girlich, M. (2022). *Tidyr: Tidy messy data*. Retrieved from
1089 <https://CRAN.R-project.org/package=tidyr>
- 1090 Wickham, H., Hester, J., & Bryan, J. (2021). *Readr: Read rectangular text data*.
1091 Retrieved from <https://CRAN.R-project.org/package=readr>

- 1092 Wickham, H., & Miller, E. (2021). *Haven: Import and export 'SPSS', 'stata' and*
1093 *'SAS' files*. Retrieved from <https://CRAN.R-project.org/package=haven>
- 1094 Wilke, C. O. (2020). *Cowplot: Streamlined plot theme and plot annotations for*
1095 *'ggplot2'*. Retrieved from <https://CRAN.R-project.org/package=cowplot>
- 1096 Xie, Y. (2015). *Dynamic documents with R and knitr* (2nd ed.). Boca Raton,
1097 Florida: Chapman; Hall/CRC. Retrieved from <https://yihui.org/knitr/>
- 1098 Zhang, H. (2021). *ARTofR: Who ever care about the [art of r] scripts?* Retrieved
1099 from <https://CRAN.R-project.org/package=ARTofR>
- 1100 Zhu, H. (2021). *kableExtra: Construct complex table with 'kable' and pipe syntax.*
1101 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	19	57.89	41.68	10.41	27.00	59.00
Novice	26	65.38	23.38	1.72	20.00	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	19	13.84	12.30	3.00	37.00	NA	NA	NA	NA	NA	NA
Novice	26	0.00	0.00	0.00	0.00	7.15	2.62	3.00	11.00	10.92	8.09

Table 3

AUDIT-C scoring table

Questions	Score				
	0	1	2	3	4
How often do you have a drink containing alcohol?	Never less	Monthly or per month	2-4 times per week	2-3 times per week	4+ times per week
How many standard drinks of alcohol do you drink on a typical day when you are drinking?	1-2 3-4	5-6	7-9	10+	
How often do you have 5 or more drinks on one occasion?	Never monthly	Less than monthly	Weekly		Daily or al-most daily

Note. Item scores are summed. Women who score 3+ and men who score 4+ are at-risk

Table 4

Scale analysis for novices' self-assessment

	N Items	M	SD	Min	Max	Skewness	Kurtosis	alpha
Classroom Management	8.00	2.93	0.26	2.50	3.38	0.24	1.77	0.37
Balance	3.00	3.17	0.56	1.67	4.00	-0.36	3.23	0.63
Presence	8.00	3.10	0.34	2.50	3.88	0.05	2.64	0.65
Natural Behavior	3.00	3.14	0.60	1.67	4.00	-0.49	2.94	0.78

Table 5

Scale analysis for experts' self-assessment

	N Items	M	SD	Min	Max	Skewness	Kurtosis	alpha
Classroom Management	8.00	3.18	0.37	2.38	3.88	0.00	2.74	0.65
Balance	3.00	3.30	0.57	2.00	4.00	-0.74	2.79	0.76
Presence	8.00	3.36	0.37	2.50	3.88	-0.61	2.76	0.76
Natural Behavior	3.00	3.44	0.45	2.67	4.00	-0.17	1.98	0.60

Table 6

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 7

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 8

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 9

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 10

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

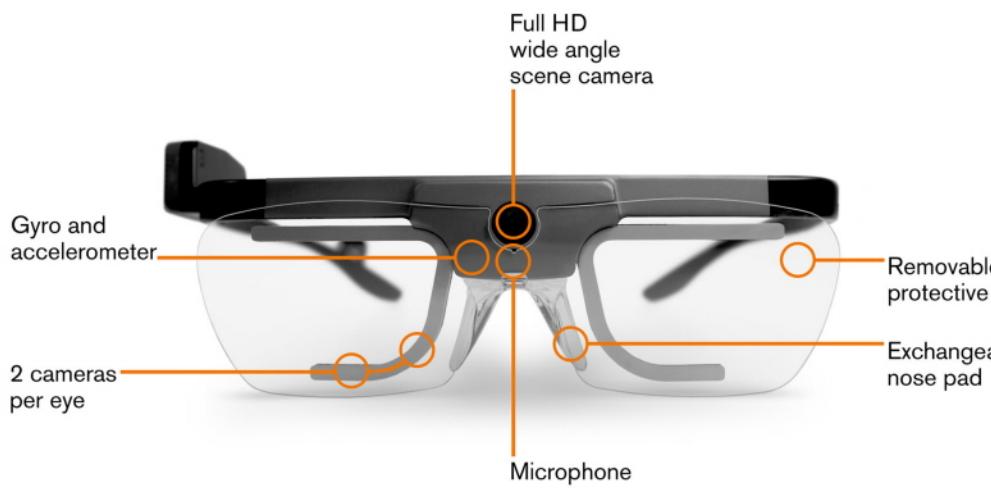


Figure 1. Tobii Pro Glasses 2; Source: <https://www.tobiipro.com/product-listing/tobii-pro-glasses-2/>

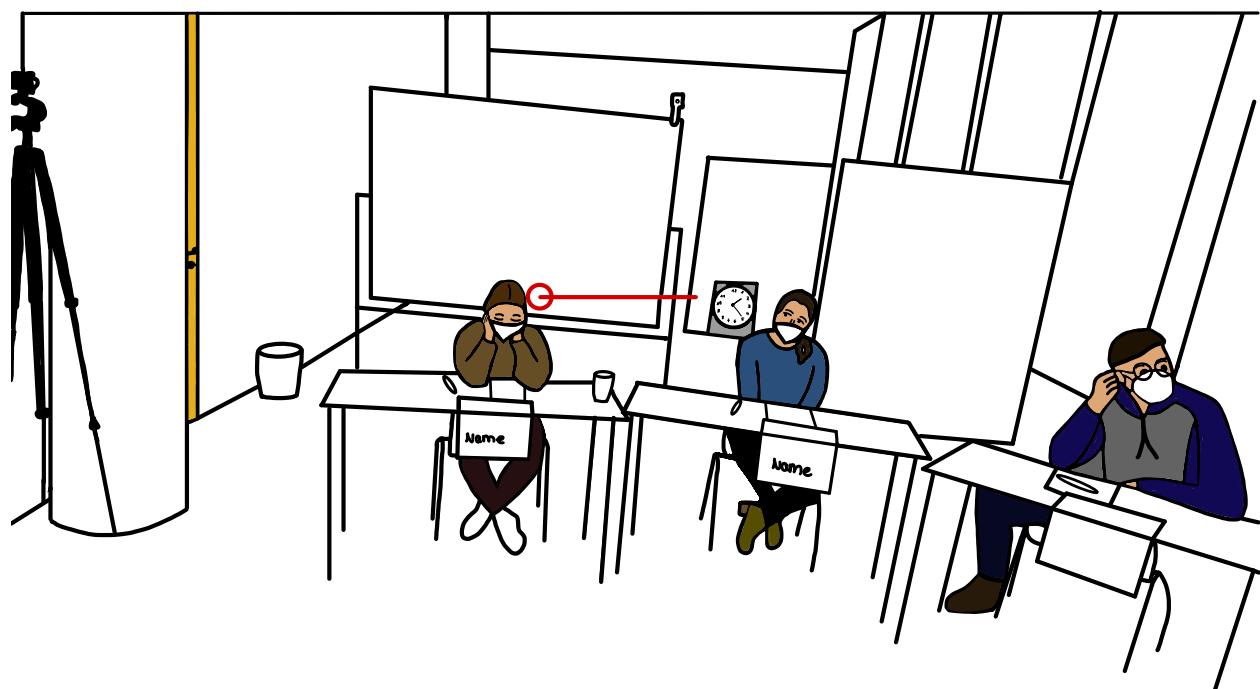


Figure 2. Teacher's Gaze Point



Figure 3. Subject and experimenter during the Stimulated Recall Interview