Classroom complexity affects student teachers’ behavior in a VR classroom

Yizhen Huang a,*, Eric Richter a, Thilo Kleickmann b, Axel Wiepke c, Dirk Richter a

a Department of Education, University of Potsdam, Germany
b Department of Research on Teaching and Teacher Education, Kiel University, Germany
c Department of Computational Science, University of Potsdam, Germany

ARTICLE INFO
Keywords:
Augmented and virtual reality
Simulations
Improving classroom teaching
Media in education
Pedagogical issues

ABSTRACT
Student teachers often struggle to keep track of everything that is happening in the classroom, and particularly to notice and respond when students cause disruptions. The complexity of the classroom environment is a potential contributing factor that has not been empirically tested. In this experimental study, we utilized a virtual reality (VR) classroom to examine whether classroom complexity affects the likelihood of student teachers noticing disruptions and how they react after noticing. Classroom complexity was operationalized as the number of disruptions and the existence of overlapping disruptions (multidimensionality) as well as the existence of parallel teaching tasks (simultaneity). Results showed that student teachers (n = 50) were less likely to notice the scripted disruptions, and also less likely to respond to the disruptions in a comprehensive and effortful manner when facing greater complexity. These results may have implications for both teacher training and the design of VR for training or research purpose. This study contributes to the field from two aspects: 1) it revealed how features of the classroom environment can affect student teachers’ noticing of and reaction to disruptions; and 2) it extends the functionality of the VR environment—from a teacher training tool to a testbed of fundamental classroom processes that are difficult to manipulate in real-life.

1. Introduction

Lively and dynamic as they may be, classrooms can also be overwhelming environments for student teachers. When a multitude of events are happening simultaneously, each one seems important, and many require quick decisions and immediate actions. For teachers in classroom situations like these, it is crucial to be aware of and act on what is really important. Student teachers often report that classroom management is their primary concern about their future work (Emmer & Stough, 2001; Flower et al., 2017). Yet, empirical evidence has shown that student teachers face great challenges in even noticing significant classroom events such as disruptions (e.g., König & Kramer, 2016; McIntyre & Foulsham, 2018; Stockero et al., 2017). The ability to notice what is happening in the classroom has been regarded as an important prerequisite for effective classroom management (Fadde & Sullivan, 2013; Gold & Holodynski, 2017). Given its significance, noticing has been investigated through the lens of different concepts in the literature: withitness (Kounin, 1970), vigilance (Parasuraman, 1986), situation awareness (Endsley, 2015; Miller, 2010), and more recently, as the constitutive element of professional vision (Schack et al., 2017; Sherin et al., 2010).
With regard to the challenges to noticing, previous observational studies pointed to a potential influencing factor: the dynamic and complicated classroom environment, or classroom complexity (Clarridge & Berliner, 1991; Phillips & Downer, 2017). Although past evidence has shown that classroom complexity may indeed cause difficulties in noticing, this intuitive association has rarely been tested in a systematic and empirical fashion. As the level of classroom complexity is difficult to manipulate in real-life settings, empirical research has been unable so far to answer the question of whether increased classroom complexity affects how teachers notice and react to disruptions. In the current study, we examined the direct influence of classroom complexity on how likely student teachers were to notice a disruption, and how they reacted to the disruptions they noticed, in a standardized virtual reality (VR) classroom. In the following sections, we will review the significance of noticing disruptions for effective classroom management, discuss the theoretical constructs underlying classroom complexity, and explore the relationship between classroom complexity and cognitive load theory.

1.1. Noticing disruptions and classroom management

From an ecological perspective on teaching, classroom management can be defined as any of the actions teachers take to use instructional time effectively and to maintain student attention, and through this, to create an environment that and facilitates both academic and social-emotional development (Emmer & Slough, 2001; Everson & Weinstein, 2006). A well-managed classroom is associated with achievement gains (Dijk et al., 2019; Freiberg et al., 2009), positive learning attitudes (Gage et al., 2018; Hung & Fan, 2014), better classroom climate (Besser et al., 2019; Ratcliff et al., 2010), and generally better school adjustment (Aldrup et al., 2018), among many other positive outcomes (Korpershoek et al., 2016). Despite teachers’ efforts to create and maintain a productive learning environment, unexpected events may occasionally occur in the classroom and interfere with instructional activities (Piwowar et al., 2018). These events have been referred to as misbehaviors, disruptions, critical situations, or secondary vectors of actions (e.g., Doyle, 2006; Gold & Holodnyski, 2017; Wolff et al., 2015). In this study, we use (student) disruptions and misbehaviors interchangeably to refer to student-initiated classroom events that may interfere with teachers’ primary program of action.

The diverse approaches teachers take in managing their classrooms are highly situation-specific (König & Kramer, 2016). Nevertheless, being able to notice disruptions when they occur is an important prerequisite for effective classroom management (Güler et al., 2020; Schack et al., 2017). The significance of the ability to notice what is happening in the classroom was addressed in Kounin’s (1970) seminal work on classroom management: In it, he found that teachers’ ability to detect problems early and communicate this awareness to their students—which he referred to as “withitness”—was associated with high student work involvement and low disturbances (Brophy, 1988). More recently, researchers who made use of videos to assess and foster teachers’ professional vision also stressed the centrality of noticing from a broader perspective (Gold & Holodnyski, 2017). According to Sherin and colleagues, teachers’ professional vision is “the ability to notice and interpret significant features in a classroom” (Sherin, 2001). Within this framework, the initial detection of relevant classroom events is the precondition for further knowledge-based reasoning (Sherin et al., 2010), prediction (Schäfer & Seidel, 2015), and action (Jacobs et al., 2010) in various aspects of teaching (van Es et al., 2017).

Noticing what is happening in the classroom is essential but often difficult for teachers, especially those just beginning their careers. Research has shown that student teachers often struggle to identify relevant events when watching classroom videos (e.g., Calandra et al., 2008; Carter et al., 1988; Stockero et al., 2017) or observing real-life classrooms (Gegenfurtner et al., 2019; Sabers et al., 1991), that they recall little details of classroom management after teaching (König & Lebens, 2012), and that based on their eye movements, they often focus on limited areas of the classroom and miss important events (Cortina et al., 2015; McIntyre & Foulsham, 2018; Stürmer et al., 2017). The significance as well as the challenge of noticing classroom situations has also been reported by studies around the globe (e.g., Huang & Li, 2012; Osmanoglu et al., 2015; Sherin, 2001; Sun, 2020; Tsukui et al., 2017).

The concept of classroom effects, which describes the influences of environment on teacher behavior and student outcomes (Ahmad et al., 2017; Ozudogru & Aksu, 2020), has been proposed to account for this difficulty in noticing (Doyle, 2006). It highlights a critical observation that has often been dismissed: besides the teacher’s own characteristics or dispositions, the situation and environment in which the teaching happens also contribute greatly to the teacher’s behavior (Vors & Kirk, 2016). Specifically, Kounin (1970, p. 1979) and Doyle (1985) argued that the innate complexity of the classroom environment makes the ability to notice crucially significant for teaching, but simultaneously challenging. These authors went on to develop an ecological approach to classroom management in which they acknowledged the association between classroom complexity and teachers’ behavior, including noticing and reacting to disruptions. In the next section, we will review how classroom complexity has been explored and dissected against the backdrop of teachers’ classroom management behavior.

1.2. The complex nature of classrooms: an ecological view

Based on close examinations of real-life classrooms, Kounin (1970) and subsequently Doyle (1977, p. 1985, 2006) identified six distinct dimensions that shape the classroom as a complex environment: multidimensionality, simultaneity, immediacy, unpredictability, publicness, and history. These dimensions of complexity capture the variety of tasks that teachers carry out in the classroom (multidimensionality); the occurrence of multiple activities at the same time (simultaneity); and the rapid and unpredictable occurrence (immediacy and unpredictability) of these activities. Additionally, classrooms are public places (publicness) in which teachers and students share experiences (history) (Doyle, 2006, p. 99). As the ecological view of classroom management takes into account all of the actions teachers take to effectively utilize instructional time and maintain students’ attention, it understands teachers’ classroom management behavior as inextricably woven into the varied levels of complexity of classroom environments, such that “the level of skill required for successful management depends on (…) the complexity of the activity being carried out” (Doyle, 1979, p. 140).
Early studies have identified some associations between the different dimensions of complexity and teachers’ ability to notice disruptions in the classroom. Student teachers have been observed, for instance, to overlook misbehaviors when the seats are scattered (Bossert, 1977), to perform “panic transitions” (shift activity prematurely) without noticing students’ reactions when facing unexpected adverse feedback from students (Arlin, 1979), to fixate on only one section of the classroom when teaching for the first time in real-life settings (Doyle, 1977), and to report high levels of distraction in crowded classrooms (Ahrentzen & Evans, 1984; Phillips & Downer, 2017). Among the many naturalistic observational studies Claridge and Berliner’s study (1991) was one of the few to explicitly control one of the dimensions of complexity by creating a staged classroom setting with multiple “confederate students” who intentionally disrupted the teacher (e.g., by being unruly, tardy, and getting up without permission); student teachers either overlooked the disruptions or expressed frustration with being unprepared to handle such situations. Findings from these early endeavors suggest that classroom complexity has an influence on student teachers’ noticing of disruptions, yet almost none of these studies operationalized or manipulated the levels of the different dimensions of classroom complexity. More critically, due to the ethnographic nature of the ecological perspective, this collection of studies did not conceptualize a systematic explanation for the possible relationships between classroom complexity and teachers’ noticing.

1.3. Cognitive load theory

To fill the conceptual gap mentioned above, the cognitive load theory has provided a framework for explaining the effect of classroom complexity on teachers’ behavior (Feldon et al., 2019; Prieto et al., 2018). Cognitive load can be considered as an index of mental effort required to accomplish a certain cognitive task (Plass et al., 2010; Salomon, 1984). Sweller (2010, 2011) further distinguished three types of cognitive load: the load imposed by the “intrinsic nature of the information” is called intrinsic cognitive load, while the load imposed by “the manner in which the information is presented” is extraneous cognitive load. Both types of cognitive load would require cognitive resources from working memory, but only the resources devoted to the intrinsic load are relevant to the task demand and therefore is referred to as germane resources or germane cognitive load (Feldon, 2007; Sweller, 1999). Levels of both intrinsic and extraneous loads are educed from the so-called element interactivity—the extent to which pieces of information required to be processed interact with one another (Plass et al., 2010). An element can be any information that needs to be processed, such as a new concept that needs to be learned. In low element interactivity scenarios, pieces of information can be processed independently and therefore impose a lower cognitive load, whereas high element interactivity scenarios consist of information that interact with each other and must be processed concurrently (Sweller, 2010). Cognitive load has typically been measured with subjective ratings, physiological measures and performance in the secondary task (e.g., Anmarkrud et al., 2019; Leppink et al., 2013; Paas et al., 2003). Essentially, cognitive load theory focuses on what underlies and connects the environmental features and observable behaviors: the human capacity for cognitive processing that is working memory, and particularly the limits of this capacity (Kalyuga & Singh, 2016; Sweller, 2020).

The notion of cognitive load is based on the premise that our primary information processing structure—working memory is limited in both capacity and duration (Ballard et al., 1995). When performing a task that requires intensive processing, the novel information from environment oftentimes impose heavy load on working memory (Sweller, 2011). Especially, when the task scenarios possess high element interactivity, the aggregated cognitive load can easily exceed the limited cognitive resources that working memory can offer, and lead to the phenomenon of cognitive overload (Bargh & Ferguson, 2000; Kirsh, 2000). This overload may greatly hinder task performance, causing, for instance, decreased accuracy and efficiency in problem solving (Adler & Benbunan-Fich, 2012; Segijn et al., 2019). Cognitive overload is also associated with difficulties in knowledge retention and transfer (Alghamdi et al., 2020; Mayer & Moreno, 2003), failures in making rapid and effective decisions (Bolisani et al., 2018), as well as feelings of fatigue and stress (Lee et al., 2016).

Returning to the classroom context, the potential influence of classroom complexity on teachers’ behaviors may also be explained by cognitive load theory. In the classroom environment, teachers are facing concurrent tasks including but not limited to navigating the classroom, monitoring students’ activities, attending to students’ responses, delivering content, employing instructional technologies, and much more. Each of these tasks demands some portion of teachers’ cognitive resources (Feldon, 2007; Hebert, 2018; Maranges et al., 2017). Furthermore, these tasks need to be processed parallelly and rapidly, raising the element interactivity of the situation even higher. Therefore, the dynamic and complicated classroom environment may impose immense cognitive demands on teachers as they are required to notice and react to disruptions rapidly while attending to all students and implementing their lesson plan simultaneously (Feldon, 2007). For teachers faced with these multiple demands, cognitive overload is highly likely to happen (Dessus et al., 2015; Hebert, 2018; Moos & Pitton, 2014).

The application of cognitive load theory has been particularly prominent in learning environment design, such as students’ learning by decreasing extraneous load imposed by the environment and increasing germane load that’s relevant to schema acquisition (e.g., Anmarkrud et al., 2019; Boulton et al., 2018; Mutlu-Bayraktar et al., 2019). Although teachers’ cognitive load imposed by the classroom environment has not yet been measured directly in the classroom, observations have indicated that student teachers’ behaviors in complicated situations resemble behaviors observed under cognitive overload. When faced with cognitive overload, people often attempt to reduce task complexity by relying on “fast and frugal” strategies (Gilovich et al., 2002). Applied to a classroom setting, this would mean that when a disruption has occurred and teachers are required to perform multiple tasks simultaneously, they would make fewer transitions and limit themselves to simpler activities such as structured seat work or turn-taking (Clark, 1988; Mchoul, 1978); they would also explain tasks and concepts in an oversimplified fashion (Duffy & McIntyre, 1982) or change their original lesson plan and leave out key material (Livingston & Borko, 1989). In addition to implying the use of fast and frugal instructional decisions, the cognitive load imposed by classroom complexity may also pose a challenge to noticing, as described above in Section 1.1.
This may include focusing attention on a small area of the classroom (McIntyre & Foulsham, 2018; Wolff et al., 2017) and failing to notice significant disruptions (Stockero et al., 2017; Stürmer et al., 2017).

In sum, the aforementioned studies present a picture of student teachers being challenged by classroom complexity and in some cases suffering consequences similar to cognitive overload, eventually resulting in difficulties in noticing significant events and in executing multiple tasks simultaneously. Yet as the key variable of classroom complexity has rarely been controlled and quantified in the research, the potential relationship between classroom complexity and student teachers’ noticing and reacting to disruptions is still not empirically established.

1.4. Virtual reality classroom as testbed

As aforementioned, the ecological view proposed that classroom features have an important impact on teacher’s classroom management behaviors, and cognitive load theory has contributed to explaining how these impacts unfold in finer details. Jointly, they inferred a potential factor accounting for student teacher’s classroom management behavior—classroom complexity. Yet to this day, the association between classroom complexity and how teachers notice and react to disruptions has still not been empirically validated and can only be derived from theoretical considerations or anecdotal observations.

The missing piece of the picture is the quantification and experimental manipulation of classroom complexity dimensions. It is immensely difficult, if not impossible, to manipulate the level of complexity in real-world classroom situations. An environment that balances the complexity of real-life teaching and the precision of experimental control is prerequisite for examining the underlying mechanism underlying teacher’s classroom management behavior. To this end, virtual reality (VR) has emerged as a promising candidate (Radianti et al., 2020; Ye et al., 2019). VR is a category of technologies that provides “synthetic, highly interactive three dimensional (3D) spatial environments that represent real or non-real situations” (Mikropoulos & Natsis, 2011). VR environment can either be immersive or non-immersive, depends on the sensory access to the real world. For instance, head-mounted displays can block visual and auditory input from reality, providing a fully immersive VR experience (Concannon et al., 2019), while non-immersive VR environments are mostly displayed on traditional flat screens (Jensen & Konradsen, 2018).

VR offers two key advantages when it comes to empirical research. First, VR creates a sense of presence through a first-person viewpoint and autonomous interactivity (Dede, 2009; Slater, 2018). Presence is “the sense of being there” (Mikropoulos, 2006). It describes the subjective impression that one is participating in a comprehensive, realistic experience, leading to the belief that the virtual world is true and real (Bowman & McMaham, 2007; Sadowski & Stanney, 2002). A strong sense of presence enables participants to perceive and act as if in reality, resulting in a high ecological validity and transferability of study results (Roberts et al., 2019). Second, VR not only provides a compelling experience that closely resembles a real-life situation but also ensures the standardization of experimental controls that are difficult to achieve in reality. All aspects of the VR environment are programmed and therefore under the total control of the experimenter (Tarr & Warren, 2002). For instance, in the VR environment, researchers can strategically manipulate the frequency and duration of a variety of situations that are unpredictable in real-life classrooms (Theelen et al., 2019). Environmental features such as classroom size, layout, student characteristics, and the number and types of disruptions can also be defined and controlled (McGarr, 2020; Ye et al., 2019).

Implementations of VR in empirical research are vast and varied, with teachers as the targeted population, the biggest application of VR in the field has been teacher education (McGarr, 2020). Over the last decade, VR has increasingly been used as a practical and realistic training environment for teachers because it is safe enough for exploration, while also being reliable enough for repeated practice (Jensen & Konradsen, 2018; Radianti et al., 2020). VR classrooms such as simSchool (Deale & Pastore, 2014), TLE TeachLive™ (Dieker et al., 2014, 2019), Mursion (Hudson et al., 2018; Kaufman & Ireland, 2016) and other similar environments have demonstrated the great prospect of VR in teacher training. The VR classrooms have been developed to teachers’ acquisition of professional competencies in areas like classroom management (Hudson et al., 2019; Lugrin et al., 2016; Ye et al., 2019), science process skills (Artun et al., 2020), leading discussions (Mikeska et al., 2019; Peterson-Ahmad et al., 2018), instructional planning (Bujdós et al., 2019), social communication like giving praise and feedback (Dawson & Lignugaris/Kraft, 2017; Neutzling et al., 2018; Theelen et al., 2019) and inclusive instruction for students with special needs (Garland, Holden, & Garland, 2015). Studies have shown positive effects of simulated teaching in VR on teacher’s motivation (Graziano, 2017), self-efficacy (Gundel et al., 2019), stress, and negative emotion (Stavroulia et al., 2019). They endorsed VR as it helps with bridging the gap between abstract theory and practical teaching skills, providing feedback with vivid examples in a cyclical fashion, and easing the performance stress experienced by student teachers with a safe place for practice and experimentation (Dalingler et al., 2020; Dieker et al., 2014; McGarr, 2020).

Despite the growing implementations of VR in educational science, most studies in the field have only employed VR as a training and assessment tool and seldom as an experimental testbed. The high ecological validity and experimental controllability of a VR environment has already presented itself as an ideal setting for studying human behavior in several branches of human subject research such as consumer behavior (Kim et al., 2020), human perception (Roberts et al., 2019; Tarr & Warren, 2002) and social behavior (Kyrliotis & Michael-Grigoriou, 2018). Multiple empirical studies regarding the influencing factor of teachers’ behavior have already attested the potential value of using VR as an experimental testbed (e.g., Heitzmann et al., 2019; Kaiser et al., 2013). Grounded in this tradition, we are broadening the usage scenario of VR in educational science by using VR as a virtual laboratory, for the sake of uncovering the environmental factors influencing student teacher’s classroom management behavior.

1.5. Aims of the current study

The current study aims at examining the effects of complexity features of classroom environment on student teachers’ likelihood of
noticing disruptions, as well as the fashion of their subsequent reactions. Our research aim is directly built on the ecological view of teacher behavior: how teachers behave is at least partly influenced by the environment in which the teaching takes place. In the present study, we are aiming to contribute to the field by empirically examining the claims from the ecological view of classroom management and by extending the functionality of VR—from a tool mostly used for training to a realistic and configurable laboratory. Therefore, we examined the impacts of classroom complexity on student teachers’ likelihood of noticing a disruption, as well as the types of reactions they displayed once they noticed the disruption, in a controlled yet realistic VR environment. We altered the level of classroom complexity with regard to the first two complexity dimensions described by Doyle (1977, 2006): multidimensionality and simultaneity. We defined the classroom complexity level as high when the environment was more multidimensional and simultaneous (see Table 1 column 3 for operationalizations). The other complexity dimensions were not included in this study as they were either constant or irrelevant in the VR classroom. **Immediacy** (classroom events take place at a rapid pace and require immediate decisions) overlaps with the **simultaneity** dimension in this study; therefore, only simultaneity was used. **Unpredictability** (classroom events are difficult to anticipate or be prepared for) was a constant, not a variable, in the VR classroom, as all the participants were new to this environment and could not anticipate the events. **History** (classes meet for multiple times over a longer period; participants share a common set of experiences, routines, and norms that influence the current lesson) was not applicable in the VR classroom as the teaching simulation only occurs once, and the “students” have no memory of the past events. **Publicness** (classrooms are public places, and the teaching is witnessed by a large group of students) is also not applicable in the VR classroom, as all the students are virtual 3D models.

We hypothesized that student teacher noticing of and reactions to disruptions are affected by the complexity level of the VR classroom. As suggested by cognitive load theory, complex classroom environments would impose high levels of cognitive load on student teachers’ processing ability (Feldon, 2007). When verging on cognitive overload, student teachers might therefore be less observant and tend toward brief and less effortful (minor) reactions as a means to preserve their limited cognitive resources in such circumstances. Therefore, we state the following hypotheses.

**Hypothesis 1.** With a higher level of complexity (i.e., high level of multidimensionality and simultaneity) in the VR classroom, student teachers will be less likely to notice the disruption.

**Hypothesis 2.** With a higher level of complexity (i.e., high level of multidimensionality and simultaneity) in the VR classroom, student teachers will be less likely to react to the disruption in comprehensive and effortful (major) ways once they have noticed the disruption.

### 2. Methods

#### 2.1. Participants

Fifty student teachers from a public German university were recruited, an adequate sample size for the current experiment design ($\alpha = 0.05$, power$^1 = 0.99$) (Faul et al., 2007; Flight & Julious, 2016; Maxwell, 2000). Of this sample, 44% ($n = 22$) were male and 56% ($n = 28$) were female, with an average age of 23.6 years. These students were studying to become teachers, and were either pursuing a bachelor’s (82.30%) or master’s degree (17.70%). The demographic of this sample is typical for students attending teacher education programs in Germany (e.g., Cortina & Thames, 2013; König et al., 2020). These participants were attending a weekly seminar on classroom management theory and practice. The seminar was a part of the teacher preparation program at this institution. VR teaching simulation constituted an important element of this seminar, and therefore all participants had experienced the VR classroom at least once during the semester.

#### 2.2. **Material and equipment**

For the current study, we developed a VR environment simulating a regular secondary school classroom in Germany. The VR classroom is a rectangular room with five rows and three columns of tables (see Fig. 1 left). Behind each of the fifteen tables sit two virtual students who vary in aspects of their physical appearance such as hair color, hair style, and clothing. Each student’s name is written on a name card on the table in front of her or him (see Fig. 1 right). There are other room features as well as ambient classroom noises that are designed to ensure the sense of immersion within this environment. A pilot study that utilized this same VR classroom had previously demonstrated that student teachers found this environment to be authentic and believable (Wiepke, Richter, Zender, & Richter, 2019).

For the hardware, we used the HTC VIVE Pro head-mounted device (VR headset) for presenting the VR classroom. It is equipped with a high-resolution display (2880 $\times$ 1600 pixels) and soundscape ensuring an immersive experience. The unique advantage of this device is the coordination between not only the head but also the body movement and the virtual view. This means the participant can walk around in the real-world while experiencing corresponding visual/auditory feedback within the virtual world.

During the experiment, the onset time, duration, location, and type of disruptions were predetermined. The virtual students appeared to be engaged with the lesson but would not actually interact with the participants. The sequence of disruptions was

---

$^1$ Post hoc power analysis for linear multiple regression with effect size 0.5, two tails and two predictors.
automatically initiated once the participant entered the simulation and would not be altered by participants’ actions. Thus, everything that happened during the simulation was standardized across the different participants within each condition. The disruptions were selected from a pool of typical disruptions that were rated by experts to be clearly not related to class content and that constituted significant interruptions to the class. They included chatting with classmates, throwing balls of paper, or punching classmates (Borko, 2016; Wolff et al., 2016).

2.3. Research design and procedures

The current study used a two-way mixed design with two factors, each corresponding to one complexity dimension: a within-subject factor of task (monitoring or instruction) denoting multidimensionality; and a between-subject factor of disruption level (high or low) entailing simultaneity.

For the first factor, the experiment session included two tasks for all participants: monitoring and instruction. The monitoring task required participants to manage the class by having students “do seat-work”, whereas the instruction task required them to also give a prepared lecture while monitoring and managing the classroom. Therefore, the instruction task was more multidimensional than the monitoring task.

For the second factor, each participant was randomly assigned to one of the two disruption levels. The low disruption level condition included six non-overlapping misbehaviors in a period of 2:45 min. The high disruption level condition contained twelve misbehaviors, some of which overlapped during the same duration of time (see Appendix for the script of disruptions). The high disruption level was more complex than the low disruption level in terms of simultaneity—there were more disruptions happening simultaneously. The number and order of misbehaviors were determined arbitrarily by an expert board consists of four researchers specialized in teacher education. The design of the disruption script was guided by several considerations: typical disruptive situations from teacher training video materials (Thiel et al., 2012) were taken into account; the frequency of disruptions should feel natural; the high and low disruption level should be clearly distinguishable.

More specifically, the stimuli (disruptions) were crossed with the within-subject factor (task), meaning that each participant underwent both tasks (instruction and monitoring) and was therefore presented the same set of disruptions in each of the tasks. The two tasks only differed in the audio instruction and both lasted for 2:45 min. Depending on the disruption level to which the participant was

Table 1
Classroom Complexity Dimensions and Their Operationalizations in the VR Classroom.

<table>
<thead>
<tr>
<th>Dimension</th>
<th>Explanation</th>
<th>Operationalization</th>
</tr>
</thead>
<tbody>
<tr>
<td>Multidimensionality</td>
<td>Various tasks and objectives coexist in the classroom.</td>
<td>The existence of parallel tasks, i.e., monitoring only (monitoring task) or monitoring plus instruction (instruction task).</td>
</tr>
<tr>
<td>Simultaneity</td>
<td>Many things happen simultaneously in the classroom.</td>
<td>The quantity of disruptions and the existence of overlapping disruptions, i.e., disruption level (high or low).</td>
</tr>
</tbody>
</table>

Note. The explanations for the complexity dimensions were adapted from Doyle (1977, 2006).

![Fig. 1. Demonstration of the VR classroom (left: room layout; right: participant’s view).](image_url)
assigned, she or he faced either twelve (high level) or six (low level) disruptions during each task (see Table 2).

To summarize and describe the steps of our experiment: the session began with the participant putting on the VR headset. The participant’s viewpoint in VR changed as the participant turned and moved his or her own body and head in real life. The participant could move around freely and interact with certain objects in the virtual classroom, but not with the virtual students. Each participant was assigned to either the high or the low disruption level condition while completing the tasks of both monitoring and instruction. The participant first practiced how to navigate in the VR classroom after listening to a prerecorded audio guide. Then the participant was prompted to start with either the instruction or monitoring task in randomized order. The task prompt stated that the participant was a substitute teacher who had to step in at the last minute and therefore did not know the students’ names or backgrounds. The participant was instructed to address the students by the names on their name cards while teaching. The participant was also instructed to “teach as if in a real classroom” and to manage the classroom as they deemed appropriate. After the first task, the audio guide gave instructions for the next task. The participant then completed the second task with the same level of classroom disruption. The whole session lasted around 15 min.

2.4. Measures

We were interested in two response variables: noticing and reacting. Therefore, two coded responses were assigned to each participant for each disruption: whether the participant noticed the disruption or not (0/1), and if so, whether the participant reacted in a major (comprehensive and effortful) way or not (0/1).

To acquire these responses, we recorded student teachers’ behaviors in the VR classroom from their first-person perspective. Participants’ behaviors were then coded independently by two trained coders who watched the videos and assigned values to each scripted disruption. The coded variables included noticing (noticed/not noticed) and reacting (minor/major reaction) (see Table 3 column 2 for definitions). The misbehavior is noticed if it’s a) in teacher’s first-person view when it occurs; and b) received subsequent reaction. For instance, if the disruption was not in the participant’s view as it happened, this event was coded as not noticed. Once the disruption was coded as noticed, the subsequent reaction was categorized as either brief and seemingly effortless and only disrupting the original course of action in a minor way (minor reaction), or as lengthy and effortful and disrupting the original course of action in a major way (major reaction) (see Table 3 column 3 for examples). The discrepancies between the coders (inter-rater reliability: 88.5% for noticing, 81.3% for reacting) were reconciled after supervised discussion and eventually compiled into one coding.

2.5. Data analysis

The questions we wanted to answer in this study were whether the level of classroom complexity affects: 1) student teachers’ likelihood of noticing the disruption (binary: noticed or not noticed); 2) student teachers’ likelihood of reacting to the disruption in a certain manner (binary: major or minor reaction). In order to model the likelihood of noticing and reacting as a function of the two complexity factors (disruption level and task), we built random intercept generalized linear mixed models (GLMM) with two fixed effects terms and two random effects terms for each response variable. GLMM is an extension of generalized linear model (GLM) and linear mixed models. GLMM is more appropriate than GLM or linear mixed models for the current dataset based on two considerations. First, each participant was measured in both tasks; therefore, it is necessary to include subject-specific parameters (random effects) to account for the potential correlations between the observations from the same participant (Molenberghs & Verbeke, 2011). Second, the responses (noticing, reacting) are binary variables that do not follow the assumed Gaussian distribution in mixed effects models. In short, GLMM is what we needed for analyzing non-Gaussian-distributed and correlated data.

In the random intercept GLMM we built, the fixed effects terms described the response variable as a function of the two explanatory variables that represent classroom complexity. On the other hand, the two random terms accounted for the correlations between the repeated measurements of participants’ responses in two tasks and for the same stimulus (disruptions). A random intercept model could include the effect of individual differences among participants by assuming a random intercept that follows a normal distribution with certain variance. By varying the intercepts from participant to participant and from disruption to disruption, we accounted for the possible correlation between the observations from the same participant and from the same disruption (Baayen et al., 2008; Zuur,

<table>
<thead>
<tr>
<th>Coding</th>
<th>Definition</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Noticed/not noticed</td>
<td>The misbehavior is noticed if it’s a) in teacher’s first-person view when it occurs; and b) received subsequent reaction.</td>
<td>When a certain student misbehaves, the student was not in teacher’s view or it is briefly in view but does not receive any subsequent reaction, then this misbehavior is not noticed. When a certain student misbehaves, the student teacher establishes eye contact with the student; physically approaches the student; stops teaching briefly and waits; calls on the student causing the disturbance by name, and quickly moves on (“James, do you have a question?”), etc.</td>
</tr>
<tr>
<td>Minor reaction</td>
<td>The reaction is brief and effortless; there is no or only a minimal interruption to teaching.</td>
<td>When a certain student misbehaves, the student teacher stops the lesson and addresses the misbehavior directly and rebukes the student at length; reminds the student of the consequences (“You now have the choice to either … or …”); follows up with the consequences mentioned (“Now, sit down”), etc.</td>
</tr>
<tr>
<td>Major reaction</td>
<td>The reaction is extended and effortful; there is a major interruption to teaching.</td>
<td></td>
</tr>
</tbody>
</table>
The above descriptions of our random intercept GLMM model can be expressed in a two-sided linear formula specified by the glmer function in the lme4 package (Bates et al., 2015) in R (R Core Team, 2019) (see Formula 1 and Formula 2). In these two formulas, the variable on the left of the ~ operator are the two response variables (noticing and reacting). On the right-hand side of the ~ operator are the terms separated by + operators. Random-effects terms are distinguished by vertical bars (|) separating expressions for design matrices from grouping factors. For both response variables, we were interested in the same set of terms: task, disruption level, the interaction between the two, and two random intercepts to account for influences from the individual participant and the individual stimulus. These two GLMM were fitted with binomial link functions using glmer, and the effect sizes were calculated with the SIMR package (Green & MacLeod, 2016) in R (R Core Team, 2019).

Formula 1.

\[ \text{noticing} \sim \text{task} \times \text{disruption level} + (1|\text{participant id}) + (1|\text{disruption id}) \]

Formula 2.

\[ \text{reacting} \sim \text{task} \times \text{disruption level} + (1|\text{participant id}) + (1|\text{disruption id}) \]

3. Results

3.1. Hypothesis 1

In the first research question, we hypothesized that a higher level of classroom complexity (high disruption level, instruction task) led to lower likelihood of student teachers noticing the disruption. GLMM results showed that both the task and disruption level had significant marginal effects on noticing ($\hat{\beta}_{\text{task}} = -0.73$, $p < .01$; $\hat{\beta}_{\text{disruption level}} = -0.63$, $p < .05$) with medium to large effect sizes (between 0.5 and 0.8) and without significant interactions (see Table 4). Specifically, the direction of effects can be seen with odds ratios: the instruction task and high disruption level led to smaller chances of predicting the probability of noticing by 0.48 and 0.53 times respectively, comparing to the monitoring task and the low disruption level. In short, we found that student teachers were less likely to notice the disruptions in the more complex situations—during the instruction task and in the high disruption level condition (see Fig. 2).

3.2. Hypothesis 2

Given the result that the two complexity factors affected the likelihood of noticing the disruption, the natural question is whether these factors were also associated with the way student teachers reacted to the disruptions once they noticed them. Our second hypothesis was that with a higher level of complexity (high disruption level, instruction task) in the VR classroom, student teachers will be less likely to react to the disruption in extended and effortful (major) ways once they noticed the disruption.

Similar with the percentage of noticing, we had coded the student teachers’ reactions to the disruption as a binary variable: once noticed, did they react to the disruption in a major way or not. The data used to test Hypothesis 2 therefore excluded cases that were coded as not noticed. Results showed that both task and disruption level had significant marginal effects on reacting ($\hat{\beta}_{\text{task}} = -0.92$, $p < .05$; $\hat{\beta}_{\text{disruption level}} = -0.89$, $p < .05$) with large effect size (above 0.8) and without significant interactions (see Table 5). The odds ratios demonstrated that compared to the monitoring task, the instruction task had smaller odds of predicting the probability of major reactions by 0.40 times, i.e., the probability of reacting in major ways during the instruction task was reduced by 60 percent comparing to the monitoring task. Compared to the low disruption level, the high disruption level condition had smaller odds of predicting the probability of major reaction by 0.41 times. To summarize, we found that similar to the likelihood to notice disruptions, the student teachers were also less likely to react in major ways in the more complex situations—during the instruction task and in the high disruption level condition (see Fig. 3).

4. Discussion

4.1. Conclusions

Stemming from the ecological view of classroom management and cognitive load theory, we were striving to answer the question: how do the complexity features of classroom environment affect student teachers’ likelihood of noticing disruptions, and what is the

Table 4

<table>
<thead>
<tr>
<th>Predictors</th>
<th>Estimate</th>
<th>Odds ratios</th>
<th>Effect size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Task (reference: monitoring)</td>
<td>-.73**</td>
<td>.48</td>
<td>-.73</td>
</tr>
<tr>
<td>Disruption level (reference: low)</td>
<td>-.63*</td>
<td>.53</td>
<td>-.63</td>
</tr>
<tr>
<td>Task * Disruption level</td>
<td>.08</td>
<td>.93</td>
<td>-.08</td>
</tr>
</tbody>
</table>

Note. * indicates $p < .05$. ** indicates $p < .01$. 

Y. Huang et al.
fashion of their subsequent reactions? We used task and disruption levels to operationalize classroom complexity (multidimensionality and simultaneity) (Doyle, 2006). By altering these features in a VR classroom and examining their impacts on student teachers’ classroom management behaviors, we found clear associations between the two: in situations that were characterized by greater classroom complexity, the student teachers were less likely to notice the disruption and also less likely to react to the disruptions they noticed in an effortful manner.

These findings provided empirical support for the observations reported in various studies around the globe (e.g., Osmanoglu et al., 2015; Wolff et al., 2017). In a complex classroom environment, student teachers may suffer from cognitive overload, causing them to overlook important events and switch to strategies that require less effort. Indeed, in one of the seminal ecological studies with careful classroom observation over a three-year practical training phase, Doyle found that student teachers developed strategies to reduce the complexity of classroom environment by “localizing attention to one region of the classroom” and “being engrossed in one activity at a time” (Doyle, 1977, p. 54). This insightful observation resonates with the idea from cognitive load theory that to reduce the cognitive load imposed by certain tasks, a common and often unintentional strategy is to ignore or deal with such tasks in less effortful fashion (Mayer & Moreno, 2003).

It is also important to note that the tendency to react to the disruption in a brief and seemingly effortless fashion in complex situations is not inherently good or bad. Some researchers have even claimed that overt interventions into disruptions could be risky because the teacher’s behavior could draw attention to the disruption and create a “ripple effect” that further derails the original instruction (Kounin & Gump, 1974). We would like to stress that no particular classroom management strategy is appropriate in all situations, with all the complexities that classroom environments impose. And yet the fact that student teachers are likely to overlook...
or be unaware of significant classroom events and to respond in ways that are affected by environmental features instead of going through a conscious and controlled decision-making process is worth more detailed examination to say the least.

4.2. Limitations

The current study can be further improved in four respects. First, we did not explicitly measure the cognitive load presumably imposed by the complexity factors. To further establish the mechanics of student teachers’ cognitive processing, we need to quantify the cognitive load each complexity factor imposes on student teachers. Subjective self-reports, physiological indicators and performance on secondary tasks are the common methods for measuring cognitive load in the literature (e.g., Anmarkrud et al., 2019). Therefore, we intend to replicate the current study with cognitive load measures in the future. Specifically, as classroom complexity features may have influenced student teachers’ noticing and reacting through the mechanism of cognitive overload, the level of cognitive load experienced by the participants may be moderating the effect of complexity features (Mutlu-Bayraktar et al., 2019).

Second, given that the current study focused on the influence of complexity features of classroom environment on student teachers’ likelihood of noticing disruptions and the fashion of their subsequent reactions, we had not inspected the effect of knowledge close enough. Therefore, a follow-up study with a control group of low-knowledge participants would help to clarify the influence of prior knowledge, in conjunction with complexity features, on their noticing and reacting to disruptions. Third, we defined noticing using the subjective interpretations of coders. Low-inference measures such as eye fixation would be more accurate and appropriate for defining whether the student teacher noticed the disruption (e.g., McIntyre et al., 2019; Seidel, Schnitzler, Kosel, Stürmer, & Holzberger, 2020). Finally, although our participants had reported strong sense of presence and generally rated the environment to be highly believable, we acknowledge the artificiality of the VR classroom and the subsequent generalizability issue, especially considering the virtual students were not responsive to teachers’ behaviors. In order to compare teacher’s behavior in and out of the VR classroom, replication of the current results using staged classroom situations in real-life or video snippets would be worth pursuing (e.g., Kleinknecht & Gröschner, 2016; König et al., 2014).

4.3. Implications

Despite the margin of improvement, the current study examined the key claim of the ecological view about classroom management: classroom complexity affects student teachers’ noticing and reacting to disruptions. This study was also among the first endeavors to extend the application of the VR environment from a training tool to an experimental testbed that combines high ecological validity with precise controllability in educational science. The results shed light on the further application of VR in understanding teachers’ behavior inside the classroom.

Many fields have already adopted VR as a virtual laboratory for eliciting and assessing human behavior (Kim et al., 2020; Roberts et al., 2019; Vasser et al., 2017). For example, cognitive psychologists have designed customizable VR labs for conducting experiments on spatial navigation (Alsbury-Nealy et al., 2020), social interaction (Parsons, 2015), attention and memory (Vasser et al., 2017). The field of educational research has begun to benefit from this new technological advancement over the last decade with the fast development in affordable consumer VR headsets such as Oculus Rift, HTC Vive, and Microsoft HoloLens (Radianti et al., 2020). Although VR has become increasingly accessible and commonplace for educational researchers (Kaufman & Ireland, 2016), its potential as a virtual laboratory has rarely been explored. We hope this experimental study will generate interest from various disciplines within the field of education that could benefit from using this unique technology as a laboratory.

The current study may also inspire the development of learning environments for student teachers in training. According to cognitive load theory, effective learning takes place only when the cognitive load does not exceed the available working memory capacity (Mutlu-Bayraktar et al., 2019; Sweller, 2020). A well-designed learning environment therefore needs to assist learners with allocating their valuable cognitive resources to deal with germane load instead of extraneous load (Kalyuga & Singh, 2016; Mayer & Moreno, 2003). As indicted by our results, student teachers tend to miss disruptions when the environmental features become complex. Building on this finding, we would propose that a learning environment designed for inexperienced new teachers should strive to help them differentiate two types of distractions: one that are not relevant for teaching and the other one needs to be recognized and addressed. For instance, a VR classroom can notify the student teachers with visual cues whenever they shift attention to a poster on the wall or birds outside the window instead of focusing on the students.

Our study may also cast light on teacher preparation and contribute to the large body of works on simulation-based teacher education (e.g., Dalinger et al., 2020; Kaufman & Ireland, 2016). Combining the past evidence (e.g., Hudson et al., 2018; McGarr, 2020) and our current findings, we would like to make a general recommendation for teacher training that focuses on repeated and conscious practice in recognizing and handling common disruptive scenarios. The simulated classroom scenarios as used in our study can be useful in helping student teachers to form pattern recognition abilities as described in the literature on teachers’ professional vision (e. g., Sherin & van Es, 2009) or situation awareness (e.g., Miller, 2010) and to develop the ability to handle disruptions in the classroom. VR classroom environments might be particularly beneficial in fostering such abilities due to the following factors. First, the programmable environment can expand classroom complexity gradually to approximate the full complexity of real classrooms (Grossman et al., 2009). Second, simulation-based learning has been found to be one of the most effective ways of acquiring complex skills in various domains (Chernikova et al., 2020). VR simulation not only inherits the benefits of computer simulation, but also provides a unique environment which is highly interactive and capable of invoking a sense of presence and immersiveness (Mikropoulos, 2006; Peterson-Ahmad et al., 2018). The realistic and interactive aspects of VR ensure better transfer of skills from the training environment to the real-world situation (Dieker et al., 2014; Jensen & Konradsen, 2018). Eventually, student teachers’ actual classroom
management practice in real life may be changed through the combined effects of these factors.

CRediT author statement

**Yizhen Huang:** Conceptualization, Formal analysis, Investigation, Data Curation, Writing-Original Draft. **Eric Richter:** Conceptualization, Investigation, Data Curation, Writing-Review & Editing. **Thilo Kleckmann:** Conceptualization, Methodology, Resources, Data Curation, Writing-Review & Editing. **Axel Wiepke:** Software, Investigation. **Dirk Richter:** Resources, Writing-Review & Editing.

Declaration of competing interest

None.

We declare that there are no known conflicts of interest associated with this research.

Appendix

### Table 6
Disruption Script for High and Low Disruption Level Conditions.

<table>
<thead>
<tr>
<th>Disruption</th>
<th>Seat number and time stamp</th>
<th>Low disruption level</th>
<th>High disruption level</th>
</tr>
</thead>
<tbody>
<tr>
<td>Playing</td>
<td>7 &amp; 8 (0:30 min)</td>
<td>22 (0:45 min)</td>
<td>3 (0:10 min)</td>
</tr>
<tr>
<td>Chatting</td>
<td>13 &amp; 14 (1:00 min)</td>
<td>13 &amp; 14 (1:00 min)</td>
<td>21 (1:10 min)</td>
</tr>
<tr>
<td>Eating</td>
<td>30 (1:10 min)</td>
<td>30 (1:10 min)</td>
<td>5 (1:40 min)</td>
</tr>
<tr>
<td>Lethargic</td>
<td>13 &amp; 14 (1:20 min)</td>
<td>13 &amp; 14 (1:20 min)</td>
<td>5 (1:40 min)</td>
</tr>
<tr>
<td>Lethargic</td>
<td>23 (2:00 min)</td>
<td>23 (2:00 min)</td>
<td>14 (1:00 min)</td>
</tr>
<tr>
<td>Hitting</td>
<td>8 (0:30 min)</td>
<td>23 &amp; 24 (2:10 min)</td>
<td>14 (1:00 min)</td>
</tr>
<tr>
<td>Hitting</td>
<td>3 (2:30 min)</td>
<td>3 (2:30 min)</td>
<td>14 (2:45 min)</td>
</tr>
<tr>
<td>Chatting</td>
<td>7 (0:30 min)</td>
<td>14 (1:00 min)</td>
<td>14 (1:00 min)</td>
</tr>
<tr>
<td>Eating</td>
<td>14 (2:45 min)</td>
<td>14 (2:45 min)</td>
<td>14 (2:45 min)</td>
</tr>
</tbody>
</table>

References


