education sciences

MDPI

Article

Assessing Complex Skills in a Virtual Reality Context: Capturing the Development of Pre-Service Teachers' Classroom Management Strategies

Jolien M. Mouw [1,*], Marjon Fokkens-Bruinsma [2] and Andrea Snippe [3]

[1] Department of Educational Sciences, GION Education/Research, University of Groningen, P.O. Box 72 9700, 9712 CP Groningen, The Netherlands
[2] Department of Teacher Education, University of Groningen, P.O. Box 72 9700, 9712 CP Groningen, The Netherlands
[3] Independent Researcher, 3891 AK Zeewolde, The Netherlands
* Correspondence: j.m.mouw@rug.nl

Abstract: Simulation-based learning environments such as virtual reality (VR) can be promising to help pre-service teachers develop kindergarten-specific classroom management strategies. However, the successful integration of VR in higher education is impeded by limitations regarding the assessment of multifaceted skills such as classroom management strategies. In this paper, we aimed to understand how different assessment instruments lead to aligning or diverging conclusions regarding pre-service teachers' VR-supported over-time development of classroom management strategies in terms of behaviors and beliefs. With a variety of self-reported and observational measures, we gauged distinct aspects of pre-service teachers' classroom management strategies conceptualized as behaviors and beliefs at two time points: before and after practicing in a VR kindergarten classroom. Most measures captured an over-time development of behaviors and beliefs—except for pre-service teachers' self-reported usage of classroom management strategies. Alignment was observed between measures gauging pre-service teachers' classroom management strategies beliefs and between most, but not all, measures of classroom management strategies behaviors: Pre-service teachers' self-reported usage of classroom management strategies do not align with the observations of strategy usage. Our findings emphasize the importance of using multiple instruments to capture, assess, and appraise VR-supported development of distinct aspects of pre-service teachers' kindergarten-specific classroom management strategies.

Keywords: classroom management strategies; kindergarten classroom; pre-service teachers; teacher training; virtual reality; skill development; skill assessment

check for updates

Academic Editor: Federico Corni

Received: 18 October 2024
Revised: 16 December 2024
Accepted: 21 December 2024
Published: 24 December 2024

Citation: Mouw, J. M., Fokkens-Bruinsma, M., & Snippe, A. (2025). Assessing Complex Skills in a Virtual Reality Context: Capturing the Development of Pre-Service Teachers' Classroom Management Strategies. *Education Sciences*, *15*(1), 2. https://doi.org/10.3390/educsci15010002

Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Simulation-Based Learning Environments for Professional Skills Training

Simulation-based learning environments such as virtual reality (VR) have the potential to support the training of professional skills in higher education (e.g., Renganayagalu et al., 2021; Theelen et al., 2019). Some studies specifically evaluate the application of VR environments in teacher education to support the development of teacher skills such as presentation techniques, emotional awareness, behavioral observations, and interpersonal competencies (e.g., Billingsley et al., 2019; Huang et al., 2021). Similar to research on VR-supported learning in general (e.g., McGrath et al., 2018), these works have not yet determined how the

over-time development of professional skills training in a VR environment is best assessed. Most studies evaluating VR environments have primarily focused on aspects such as user experiences of system-based performances—often related to interactivity, presence, and immersion—usability requirements, and technology acceptance (e.g., Lugrin et al., 2016) or pertain to stand-alone VR-learning experiences instead of evaluating over-time development of complex professional skills that are often multidimensional in nature (i.e., requiring the integration of content knowledge, interpretation, and practices). We posit that, when evaluating the development of such multifaceted skills in VR training contexts, a variety of measurement instruments should be considered to overcome validity and/or reliability issues reported in previous studies appraising the potential of VR-learning environments. In this paper, we aim to illustrate the importance of using different assessment instruments if the goal is to understand how a VR kindergarten classroom can support pre-service teachers in developing classroom management strategies. We will reflect on whether and how different instruments may lead to diverging conclusions regarding CMS development and will discuss how triangulation can contribute to evaluating the potential of VR for the acquisition of complex and multidimensional professional skills such as CMS.

1.1. Virtual Reality

Virtual reality is a high-end computer-generated three-dimensional simulation technology that affords users to engage in realistic, interactive, and immersive multi-sensory experiences through a head-mounted display (e.g., Burdea & Coiffet, 2003). Recently, the use of VR in higher education has rapidly increased. Partially because of technological advances and the availability of more affordable technology but also because the affordances of simulation-based training environments have the potential for training complex professional skills (e.g., Jensen & Konradsen, 2018; Radianti et al., 2020; Renganayagalu et al., 2021). VR allows students to learn new skills in a safe, controlled, structured—yet highly authentic—environment, encouraging active exploration and interaction (Pantelidis, 2009; Renganayagalu et al., 2021), and enables supervised correction, repetition of situations, and failure without consequences (e.g., Jensen & Konradsen, 2018). Recent meta-analyses found that—compared to screen-based simulations (such as mouse-controlled games)—VR-learning environments are more effective in terms of student learning outcomes and skill development (e.g., Chernikova et al., 2020). Thus, VR can offer a realistic, safe, and powerful learning environment for professional skill development, for example, for pre-service teachers to practice and develop classroom management strategies (Mouw et al., 2020; Mouw & Fokkens-Bruinsma, 2022).

1.2. Classroom Management Strategies

In this study, we focus on pre-service teachers' classroom management strategies trained in a VR-learning environment. On a daily basis, teachers implement classroom management strategies (from here on: CMS) to respond to and manage students' behaviors, helping them establish a positive and orderly learning environment supporting teaching and learning processes (e.g., Emmer & Stough, 2001; Flower et al., 2017; Korpershoek et al., 2016, 2022; Simonsen et al., 2008). Three broad categories of CMS can be distinguished (e.g., Borich, 2011; Ritz et al., 2014):

- Anticipation strategies are aimed at preventing problems and minimizing the chance of disruptive behavior(s) occurring,
- deflection strategies are focused on redirecting (precursors of) disruptive behaviors, allowing students to correct themselves, and
- reaction strategies are implemented immediately after disruptions or misbehaviors emerge to stop and/or discipline these behaviors.

Thus, one aspect of CMS concerns teacher behavior, as CMS refers to the actions teachers take to manage student behavior. However, in this study, we conceptualize CMS not only in terms of (pre-service) teachers' CMS behaviors but also their CMS beliefs. This distinct aspect of CMS pertains to pre-service teachers' beliefs regarding their ability to effectively implement CMS (i.e., self-efficacy towards classroom management; Tschannen-Moran & Woolfolk-Hoy, 2001). Less efficacious pre-service teachers tend to use fewer effective strategies and, as a result, are likely to experience more disturbances in the classroom and perceive these disturbances as more stressful (Dicke et al., 2014). This is problematic, as negative experiences with misbehaviors and low levels of self-efficacy are key factors in teachers' well-being and attrition, as CM-related self-efficacy affects whether classroom disturbances lead to emotional exhaustion (Brown & Wynn, 2007; Dicke et al., 2014).

1.3. Developing Classroom Management Strategies for Kindergarten Education

Adequate management of student behavior requires active observation of (potentially) disturbing behaviors so that teachers can respond to these behaviors in a timely manner. However, selecting and applying CMS is one of the foremost concerns in everyday teaching practices, especially for pre-service teachers (Borich, 2011; Eisenman et al., 2015). They find it difficult to determine which specific anticipative, deflective, or reactive strategy is most effective when addressing certain behaviors or to recognize when it is time to switch to another CMS category (e.g., Emmer & Stough, 2001; Putman, 2009). CMS can be even more challenging when learning is situated in a dynamic setting involving young children, such as a kindergarten classroom (Carter & Doyle, 2006; Ritz et al., 2014). On the one hand, large developmental differences between four- and six-year-old children (This study took place in the Netherlands, where compulsory education starts at the age of five. However, most children enroll in primary school as soon as they are eligible, which is at the age of four. These four-year-olds start in "Group 1", i.e., the first year of primary education. The majority of Dutch kindergarten classrooms are heterogeneous in age as they include Group 1 and Group 2 (i.e., years 1 and 2, corresponding to pre-K and kindergarten, respectively), with attending children commonly aged between four and six.) require flexibility in teaching approaches and attunement of CMS to children's specific developmental and social-emotional needs. On the other hand, kindergarten teachers must adequately and timely correct (potentially) undesirable behaviors when dealing with easily distracted kindergartners whose (self-)regulative and inhibitive skills are not yet matured (e.g., Dobbs-Oates et al., 2011; Ritz et al., 2014).

Ideally, future teachers are equipped with effective kindergarten-specific CMS during their pre-service training—preferably during a teaching placement (i.e., internship). However, curriculum-based pre-service CMS training is often insufficient and generally lacks integration, adequacy, and effectiveness (e.g., Christofferson & Sullivan, 2015; Emmer & Stough, 2001; Jones, 2006). Pre-service teachers experience difficulties in translating more general classroom management theories (which are most often validated in upper-elementary or secondary classrooms) into daily kindergarten practices and specific CMS (e.g., Carter & Doyle, 2006; Flower et al., 2017). Moreover, pre-service teachers feel that their kindergarten internship does not provide sufficient opportunities for practicing and developing effective kindergarten-specific CMS or putting a variety of CMS into practice. Pre-service teachers tend to take over their mentor's strategies (Bauml, 2011; Zeichner & Tabachnick, 1981) and sometimes feel obliged to implement specific CMS they are highly uncomfortable with (for example, gaining attention by singing). Even if pre-service teachers are able to translate general CMS theories into kindergarten teaching practices and feel that they may diverge from their mentor's practices, they are cautious in experimenting with CMS to better understand the reciprocity between CMS and specific behaviors as this

could negatively impact the teacher-student relationship (Van Tartwijk et al., 2011). Given these impeding factors, it is essential to provide pre-service teachers with more in-training opportunities to develop a variety of kindergarten-specific CMS.

An alternative to real-world, in-vivo training of complex professional skills like CMS can be offered by simulation-based learning environments (Chernikova et al., 2020; McGarr, 2021; Van der Want & Visscher, 2024), such as a VR kindergarten classroom (see Mouw & Fokkens-Bruinsma, 2022). A VR kindergarten application provides pre-service teachers the unique opportunity to safely experiment with different kindergarten-specific CMS as this immersive environment allows pre-service teachers to freely interact with the kindergarteners while responding to a diverse range of pre-programmed (antecedents of) disruptive behaviors varying in the degree of distortion, complexity, and aggravation levels. Moreover, a VR kindergarten classroom allows for the rerun of complex situations, allowing pre-service teachers to experiment with different CMSs and better understand the effect of each CMS on specific behaviors. Being able to practice with a variety of behaviors and different CMS likely improves pre-service teachers' CMS and might help them feel more confident and efficacious regarding their CMS (e.g., self-efficacy).

1.4. Present Study: A Focus on Evaluation

Notwithstanding the potential of VR for creating a motivating and engaging realistic experience, far too little attention has been paid to a fundamental aspect of professional skill development, namely assessment in a VR context. So far, researchers have not yet determined how complex skills trained in a VR environment are best assessed. In general, most studies evaluating the effectiveness of VR environments have primarily focused on user experiences of system-based performances (e.g., interactivity, presence, and immersion), usability requirements, cybersickness, and/or technology acceptance (e.g., Lugrin et al., 2016), or pertain to stand-alone VR-learning experiences (i.e., not didactically embedded in a course). Even if specific learning outcomes are targeted, it seems that adequate assessment of these learning outcomes is a complicated task as most VR studies (1) do not specify which methods were used to evaluate learning outcomes; (2) predominantly use either knowledge tests or self-report questionnaires; (3) use instruments with questionable validity; or (4) tend to overly rely on computer-generated data (e.g., Carruth, 2017; Jensen & Konradsen, 2018; Lee & Wong, 2008; Radianti et al., 2020; Renganayagalu et al., 2021).

These unidimensional and stand-alone assessment methods that are commonly used in VR research can be adequate for capturing mastery of routine-based learning objectives (such as assembly tasks for which a single correct solution can be pre-specified). However, they are oversimplified proxies when evaluating the over-time development of complex and multifaceted skills such as CMS. We posit that to understand the potential of simulation-based learning environments such as VR to support the learning and development of multidimensional professional skills, it can be conducive to consider a variety of instruments and assessment methods applied on multiple occasions to overcome issues in the way VR-supported professional skills development is currently assessed. Moreover, assessment methods grounded in and aligned with established instructional theories and practices are needed (e.g., Radianti et al., 2020), which may warrant adaptation of existing assessment methods for use in a VR context, for example, by implementing standard evaluation methods such as reflective questions or repeated instructor-based classroom observations that are frequently used in teacher training programs to assess pre-service teachers' in-vivo teaching and classroom management skills (Poulou et al., 2019). Using such repeated observations in addition to VR-generated (timed) responses and pre-service teachers' self-reported mastery and usage of CMS at multiple time points enables mapping (over-time development of) distinct facets of CMS. It could very well

be that some facets of CMS do develop over time, whereas other aspects may be more stable, such as a multifaceted, multi-instrumental, and over-time assessment approach may allow us to more adequately capture and evaluate the development of distinct facets of pre-service teachers' kindergarten-specific CMS, which could potentially help to gain a more complete picture of CMS development. Therefore, a reflection on the degree to which these different instruments align could further our understanding of the potential of VR to support pre-service teachers' professional learning. To this end, in this paper, we aim to answer the question: To what degree do different assessment instruments lead to aligning or diverging conclusions regarding VR-supported over-time development of pre-service teachers' classroom management strategies?

We will first evaluate pre-service teachers' over-time CMS development through different measures and methods supporting a multifaceted reflection of CMS. In line with our conceptualization of CMS, we will reflect on pre-service teachers' development of both the behavioral and beliefs aspects of CMS, which are captured by different instruments (i.e., self-report questionnaires and behavioral observations). Using these data sources as input, we will then reflect on the extent to which the outcomes of these instruments align in terms of pre-service teachers' over-time CMS development.

2. Materials and Methods

2.1. Design, Participants, and Procedure

A teaching-innovation grant enabled us to develop an immersive VR kindergarten classroom in which second-year pre-service teachers of our university's Academic Teacher Training Program could practice with a variety of (potentially) disturbing behaviors and safely experiment with different CMS. The VR environment was part of the 'Keeping Order in the (Virtual) Classroom' curriculum that we implemented in an Educational and Developmental Psychology course. In this course, theoretical perspectives on classroom management and young children's social-emotional and cognitive development were discussed. Parallel to this course, pre-service teachers are based at a kindergarten internship. The goal of the 'Keeping Order' curriculum was to support pre-service teachers in developing kindergarten-specific CMS and bridging theory and practice. The curriculum involved various activities in which 46 s-year pre-service teachers participated from September 2020 to January 2021. All pre-service teachers attended a theoretical lecture on classroom management (led by the first author) and then filled out the pre-measurement questionnaire in September.

For the VR sessions, we semi-randomly (e.g., after taking into account class-based schedules) assigned half of the pre-service teachers enrolled in our course to the VR group. Pre-service teachers who were not allocated to this VR condition were offered the opportunity to experience the VR kindergarten classroom upon completion of their internship. This was performed because we could not host all students due to COVID-19' restrictions. Pre-service teachers in the VR condition were placed in small groups of 3–4 students and visited our university's VR lab twice to practice their CMS in an immersive kindergarten classroom (Mouw & Fokkens-Bruinsma, 2022), once in the beginning of October and once in December. In between the VR sessions, all pre-service teachers who enrolled in the course collaboratively reflected (in groups of approximately 10) on the effectiveness of the CMS that was used by pre-service teachers in the VR group during the first VR session. The goal of these reflection sessions was to support all pre-service teachers in translating classroom management theories to everyday teaching practices: the CMS used in VR was used as an example and sparked reflection on the degree to which theories can(not) be easily transferred to a dynamic kindergarten classroom setting. Pre-service teachers filled out a post-questionnaire at the end of the semester, either immediately

after the second VR session (VR condition) or after finishing their internship (all others). As we are interested in the development of CMS in terms of behaviors (e.g., during the VR sessions) and beliefs, we focus on the data of 24 pre-service teachers (21 women, M_{age} = 20.14, age range 18.75–21.83) who consented to participate in the VR group and completed all assigned activities.

Virtual Reality Sessions

The pre-service teachers in the VR condition prepared for the small-group VR sessions by memorizing the names of all 20 virtual kindergartners as provided in a 'classroom facebook'. In addition, they were instructed to prepare a 5-min whole-group circle-time (i.e., busy-picture) activity, which is common practice in Dutch kindergarten education. At the beginning of the first VR session, participating pre-service teachers received instruction on how to interact with the kindergartners. Moreover, they were instructed on how to navigate (walk) in the virtual environment. This was particularly important as some CMS require movement (i.e., kneeling in the vicinity of a kindergartener). We set up the VR environment and the physical space in which the sessions were placed in such a way that the pre-service teachers were able to walk around the circle. They were informed that someone would help them out with the cable as soon as they would stand up so that they would not in any way be constrained by the 5 m long cable. Last, they were given time to familiarize themselves with visual, auditory, and behavioral aspects of the environment and could try out some CMS before the actual practice session started.

During the VR sessions, pre-service teachers took turns and individually led a whole-class busy-picture activity while managing the (potentially disturbing) behaviors of 20 virtual kindergartners. Other group members could observe the session live through a large monitor and note down their feedback.

The behaviors of the kindergartners were activated by one of the instructors (i.e., the authors). To offer each PST within a small group equally complex but diverse behaviors, we worked with four behavioral scripts specifying which behaviors or scenarios (i.e., a pre-determined set of automatically succeeding behaviors, for example, activation of whole-class unrest after a fire truck drives by) should be activated and how many kindergartners are involved. The four scripts were comparable in terms of complexity level and type of behaviors (e.g., looking to the right in Script 1, to the left in Script 2, to the ceiling in Script 3, and backward in Script 4; see Appendix A, for an overview of the four scripts). The scripts were matched between groups and across sessions. In the second session, pre-service teachers were offered one of the other three scripts to prevent a retest effect and to allow them to practice with different behaviors. Scripts in Session 2 generally include more complex behaviors in terms of intensity and/or the number of avatars in which behaviors were activated simultaneously (see Appendix A). In addition to the fixed scripts, instructors could activate behaviors based on lesson content or pre-service teachers' reactions. For example, when a pre-service teacher starts talking about the sirens of the emergency services (visible in the busy picture), the instructor could choose to activate 'beeping sounds' to simulate kindergarteners mimicking the sirens. A VR practice session was ended once all behaviors and scenarios of a script were managed or after 8 min of practicing to prevent overload. Then, the pre-service teacher received feedback from the instructors (and/or their peers when present (As the semester progressed, more strict COVID-19 measures were put in place. Consequently, we had to switch to individual practice sessions instead of group-based sessions in December (i.e., VR session 2).)). After receiving feedback, pre-service teachers rotated, and another group member practiced in the VR kindergarten classroom.

2.2. Measures

Throughout the semester, we used different instruments and measurement methods to gauge pre-service teachers' CMS behaviors and self-efficacy beliefs. A table with the correlations between all measures at both time points can be found in Appendix B.

2.2.1. CMS Behavior

To evaluate (changes in) CMS behavior, we used two data sources: (1) quantitative questionnaire data to gauge pre-service teachers' perceptions of their CMS usage as measured before and after practicing in the VR classroom, and (2) observational data (i.e., screen recordings) to reflect on their actual use of different types of CMS (categories) and the adequacy and timeliness of these CMS during two VR sessions.

For pre-service teachers' perceptions of CMS use, we used a researcher-constructed questionnaire that was administered prior to the first and after the second VR session. This instrument is based on the works of Borich (2011) and Kounin (1970) and entails 17 items that correspond to specific anticipation (i.e., preventing misbehavior, with-fitness), deflection (i.e., changing behavior), or reaction (i.e., disciplining behavior) strategies (see Table 1), that are thought to positively impact students' task-directed and compliant behaviors. Pre-service teachers indicated on a 5-point Likert scale (1 = never to 5 = always 1) how often they apply each CMS when teaching kindergarten. An exemplary item is: "How often do you implement 'praising' as a CMS?" Even though the CMS included in this questionnaire is theoretically aligned with anticipation, deflection, and reaction strategies, this trichotomy was not supported by statistical evidence (i.e., scale reliability), possibly because only three reactive items were included. Therefore, we calculated overall scale scores for the pre- ($\alpha = 0.83$) and post-measurement ($\alpha = 0.61$) based on all 17 items.

Table 1. Coding Scheme Observed Anticipation, Deflection, Reaction, and Ineffective Kindergarten-Specific Classroom Management Strategies.

Classroom Management Strategy	Description	(Observable) Examples in a VR Context
Anticipation strategies		
With-itness	Having eyes at the back of your head. Awareness of what happens in the classroom by observing behaviors emerging in different parts of the classroom. Knowing who is involved.	"Quincy, please stop doing that. You too, Aaron." "I saw that Isaak!"
Multitasking/overlapping	The ability to simultaneously guide and monitor (or switch between) multiple activities, situations, or kindergartners.	(Non-verbally) directing a kindergartner to the bathroom while continuing the busy-picture activity.
Engaging whole group	Actively engaging all kindergartners (physically or cognitively), keeping them on-task. For example, by whole-group questioning, letting kindergartners respond through gesturing or hand raising, interactively prompting participation in whole-class routines/activities.	"Raise your hand if you've ever visited a farm." "Who has ever seen an ambulance?" "Can you all moo like a cow?" "1, 2, 3, place your hands on your knee and look at me." "Let's see if there is hospital on this busy picture."
Flow (in transitions)	Ensuring smooth transitions between activities. Fluently recapturing focus after disruptions occurred, for example by integrating whatever just happened in the ongoing activity. Making sure all are ready to participate without taking too long to start.	"What a coincidence: a fire truck passes by just when we're talking about the emergency services! Let's see if we can find one on the busy picture." "In a few seconds, I will ask the first question so be ready."

Table 1. *Cont.*

Classroom Management Strategy	Description	(Observable) Examples in a VR Context
Scanning	Actively scanning back and forth to (1) let children know you have got eyes on them or (2) size up the significance of potential (mis)behaviors.	PST scans circle from left to right (and back) while talking. Often just before, or immediately after, turning to the busy picture on the board.
Picking up the pace	Speeding up instruction or skipping parts to maintain engagement and interest.	"I'll speed up a bit, I know you've been listening for quite some time now."
Boost interest	Spark interest and engagement by raising/lowering voice tone or volume, being enthusiastic, moving to different parts of the room (e.g., towards table or board), introducing new objects or in-between activities, directing attention to interesting details, using gestures illustrating movement.	"I'm going to tell you a secret." "(whispering) "Pssst. . . Have you seen this?" "I brought something to show you! Look at the table, what do you see?"
Removing distraction/temptation	Removing objects that could distract individuals' or group's attention before misbehavior occurs.	"Hayley, please hand over whatever you have in your pocket."
Adjusting seating arrangement	Rearranging classroom seating in advance to minimize chances of disruptive situations occurring.	"Sam, it's better if you sit next to me." "Ann and Eve, move your chairs apart so that you can't hit each other anymore like you did yesterday."
Deflection strategies		
Proximity	Walking to, or standing/kneeling near a kindergartner to deflect misbehavior. Often—but not necessarily—accompanied by putting a hand on a shoulder.	Movement towards a target kindergartner; standing or kneeling in vicinity.
Eye-contact	Communicating a warning by making eye contact with target kindergartner, often combined with raising an eyebrow, tilting the head, or short verbal utterances (e.g., a name).	"Oona, I'm looking directly at you." "Terry, look at me." "Jesse, eyes up here please."
Prompting	Reminding kindergartners of class rules and/or active participation in ongoing activities.	"Kate, we're not doing that right now, please pay attention." "Shhh. Do not speak out of turn. Raise your hand if you want to share something." "Remember our class rule: When I'm talking, you are not."
Name dropping	Integrating a target kindergartner's name into an explanation or story.	"Over here I see a rocket, Felix, that is about to launch." "Now, if Kayleigh wants to become a doctor. . ."
Praising	Acknowledging peers' desired or appropriate behaviors by giving praise (peer recognition). Praising target kindergartners who befittingly corrected themselves.	"Look at how neatly Matteo sits on his chair! I can see he is ready to start!" "Leon, I really appreciate you raising your hand this time."
Reaction strategies		
Warning	Giving (implicit or explicit) verbal warnings to correct disruptive behaviors. Signaling a consequence will follow if behavior continues. Includes counting back.	"I don't want to give any warnings. . ." "I'm going to count to three. . ." "Rose, I'm warning you. Stop that right now."

Table 1. *Cont.*

Classroom Management Strategy	Description	(Observable) Examples in a VR Context
Loss of privileges	Signaling a kindergartner will lose certain privileges if behavior continues.	"If you don't stop that right now, you will have to stay inside during break time."
Time-out or removal	Signaling a kindergartner that he/she is no longer allowed to participate in the activity and, for example, must go to the hallway or remove his/her chair from the circle if behavior continues.	"Stop that right now, or you will have to sit at your table."
Ineffective strategies		
Negative with-itness	Not aware of who is involved. Addressing other kindergartners than the one(s) misbehaving. For example, reprimanding those sitting at the other side of the room.	"Who's making these irritating noises?"
Negative flow	Starting the busy-picture activity even though some kindergartners are not paying attention/listening, are noisy, or misbehave. Waiting too long to start or resume activity. For example, repeatedly prompting one kindergartner even though most of the others (or all) obey.	"I will just wait until everyone sits straight and listens." "I want all eyes up on the board. I said eyes up here. Eyes on the board, please. I thought I asked you all to pay attention. David, eyes up here, please."
Ignoring/not observing	Either (purposefully) ignoring or not observing disruptions and/or misbehaviors.	Whenever the instructor deactivated a behavior to which a pre-service teacher did not respond.
Teacher talks	Utterances used to respond to or follow-up on (disturbing) behaviors that cannot be classified as specific CMS. Includes giving turns to individual kindergartners or asking for an answer for instructional purposes.	"Pablo, do you have a question?" "Good morning all." "Elif, what do you see?"

Behavioral observations during the VR sessions were the second source of data. During each VR session, two types of video recordings were made. The first recording captures the timed learning opportunities created by the instructor (Figure 1, left panel), whereas the second captures what the PST sees, says, and does (HMD view, right panel). The two videos were merged into one synchronized file, enabling us to observe and evaluate the type, adequacy, and timeliness of CMS our pre-service teachers used in each session.



Figure 1. Synchronized recordings of the instructor module and the HMD. Left panel: Visualization of the instructor module used to operate the virtual kindergarten avatars. Right panel: HMD view (i.e., what the PST sees through the head-mounted display).

Observed Type of CMS

To evaluate which CMS categories pre-service teachers mostly used to manage (potentially) disruptive behaviors during both sessions, we coded all observed strategies using the coding scheme presented in Table 1. The first 17 CMS were included as a priori as these correspond to the CMS surveyed in the pre- and post-test questionnaire. These CMS are subsumed under anticipation, deflection, or reaction CMS categories (Borich, 2011; Kounin, 1970). However, some of these strategies cannot be (easily) implemented in the VR environment. For example, a kindergartner cannot be sent to the hall as movement is restricted within the classroom. Therefore, we listed VR-specific approximations (including gestures) of each CMS in terms of observable teacher behaviors that can take place in the VR classroom (third column, Table 1). A fourth category, comprising three CMS (i.e., negative with-fitness, negative flow, and ignoring/observing behaviors), was added to compile information on the usage of frequently implemented yet ineffective strategies. These strategies can be considered as the inverse of anticipation strategies. Last, we added teacher talks to code strategies that do not fall under any of the other categories, bringing the total of to-be-observed CMS to 21.

The first and third authors familiarized themselves with this coding scheme by individually coding three videos. Possible discrepancies in coding were discussed to finetune the behavioral descriptions of the coding scheme. Subsequently, they coded all recordings randomly, and the inter-rater reliability was calculated on approximately 10% of the videos. The resulting kappa ($\kappa = 0.86$) suggests a sufficient degree of inter-rater agreement (McHugh, 2012).

As we wanted to compare the type(s) of CMS categories (e.g., anticipation, deflection, reaction, and ineffective) pre-service teachers used during the two VR sessions, we first calculated a ratio in which each CMS was used as a function of activated behaviors (e.g., the absolute frequency in which each strategy was used divided by the number of behaviors that needed to be managed). We used these corrected ratios instead of percentages (i.e., the absolute frequency of each CMS divided by the total number of CMS used) to prevent collinearity, which can lead to biased standard errors and inconsistent *p*-values (Vatcheva et al., 2016). We then calculated four compound ratio scores by adding up the ratio scores of all the CMS subsumed under each category.

Adequacy and Timeliness of CMS

As efficient teachers observe more behaviors and actual observation of student behavior is a prerequisite for responding, we operationalized the adequacy of CMS behavior as a ratio of unobserved (or purposefully ignored) behaviors divided by the total number of behaviors activated by the instructor. Using a ratio score instead of absolute frequencies allowed us to control for variations between students and sessions with regard to the number of behaviors that needed to be managed.

To evaluate the timeliness of CMS, we calculated how long it took for pre-service teachers to respond to each activated behavior. This was solely performed by the first author to prevent deviations caused by different time- and/or program settings. She used Microsoft Excel to note the exact time (in minutes, seconds, milliseconds) in which activated behaviors became visible or audible in the HMD view. Once a PST responded (either verbally or nonverbally) to a specific behavior, a deactivation time was noted. For each behavior, a reaction time was calculated by subtracting the activation time from the deactivation time. For each PST, an average reaction time was calculated per session.

2.2.2. CMS Beliefs

To capture (over-time development of) pre-service teachers' CMS beliefs towards two distinct yet complementary aspects of CMS, we administered a self-report questionnaire prior to the first and after the second VR session. The first is the Efficacy for Classroom Management scale (short form Teacher's Sense of Efficacy Scale; Tschannen-Moran & Woolfolk-Hoy, 2001) to measure pre-service teachers' beliefs regarding their ability to effectively implement classroom management in general. Pre-service teachers indicated on a 9-point Likert scale (1 = not at all to 9 = completely) the extent to which they believe they are competent in four aspects of classroom management. An example item is "How well can you establish a classroom management system with each group of students?" In the current sample, the averaged CM-scale proved sufficiently reliable (pre-test, $\alpha = 0.76$; post-test, $\alpha = 0.88$).

The second instrument was similar to the researcher-constructed questionnaire used to capture pre-service teachers' self-reportage usage of 17 distinct CMS-based (Borich, 2011; Kounin, 1970). To measure pre-service teachers' CMS-competency beliefs, pre-service teachers indicated on a 5-point Likert scale (1 = fully disagree to 5 = fully agree) the extent to which they felt competent in each of the 17 CMS. An exemplary item is "I feel competent in engaging the whole group." In accordance with the questionnaire tapping into pre-service teachers' reported use of these CMS, we considered CMS competence as a unidimensional measure. The averaged scales for the pre- and post-test proved sufficiently reliable (pre-test, $\alpha = 0.70$; post-test, $\alpha = 0.72$).

3. Results

3.1. CMS Behaviors

A *t*-test on normally distributed data showed that pre-service teachers' averaged reported usage of the 17 CMS did not change over time, as the pre-test ($M = 3.68$, $SD = 0.41$) and post-test ($M = 3.75$, $SD = 0.29$) scale scores do not differ significantly, $t(23) = -0.82$, $p = 0.42$, $d = 0.17$, 95% CI [−0.57, 0.24].

The compound-ratio scores that were derived from the observational data were submitted to a 4 (CMS category: anticipation, deflection, reaction, and ineffective) × 2 (Session: Session 1, Session 2) repeated measures analysis of variance (RM-ANOVA). The two significant main effects (Category and Session) were qualified by a significant Category × Session interaction, $F(3, 23) = 11.50$, $MSE = 0.13$, $p < 0.001$, $\eta_p^2 = 0.33$. This interaction effect is of particular interest as we wanted to determine if pre-service teachers' usage of each of the CMS categories (i.e., the four types of CMS) changed over time. To understand the nature of these changes, we ran follow-up paired *t*-tests for each CMS category (Table 2). These between-sessions comparisons suggest an increase in the ratio of anticipation, deflection, and reaction strategies with respectively 0.42 ($SD = 0.44$), 0.54 ($SD = 0.67$), and 0.08 ($SD = 0.12$) points change across the sessions, whereas the overall usage of ineffective CMS decreased with 0.08 points ($SD = 0.21$).

Table 2. Means, Standard Deviations, and *t*-test Statistics for CMS categories.

	Session 1		Session 2		*t*(23)	*p*-Value	Cohen's *d*
Category	**M (SD)**	**Range**	**M (SD)**	**Range**			
Anticipation	1.01 (0.51)	0.15–2.50	1.43 (0.56)	0.45–2.76	−4.66	0.000	−0.95
Deflection	1.60 (0.46)	0.83–2.69	2.13 (0.47)	1.10–2.93	−3.94	0.001	−0.80
Reaction	0.01 (0.02)	0.00–0.10	0.09 (0.13)	0.00–0.57	−3.34	0.002	−0.68
Ineffective	0.27 (0.15)	0.05–0.58	0.20 (0.16)	0.00–0.73	1.76	0.045	0.36

Two sets of paired-sample *t*-tests were used to evaluate the adequacy and timeliness of pre-service teachers' CMS usage and possible changes over time. A Shapiro-Wilk test indicated a non-normal distribution for the ratio of ignored or missed behaviors in Session 2 that could not be solved by deleting outlier values or transforming the data. Therefore, we ran a Wilcoxon-ranked test that showed that the ratio of ignored or missed behaviors was lower in Session 2 (Mdn = 7.12%) as compared to Session 1 (Mdn = 17.42%), $z = -3.14$, $p = 0.002$, $r = -0.64$, implying an increase in the adequacy of CMS behavior. Non-normality was also observed for pre-service teachers' reaction times in Session 1. After deleting two outliers, a *t*-test could be performed. We found that pre-service teachers were significantly faster in the second session, $t(22) = 3.35$, $p = 0.003$, $d = 0.72$, 95% CI [0.24, 1.18], with averaged reaction times of 7.29 s (SD = 2.04) in Session 2 and 9.54 s (SD = 3.73) in Session 1.

3.2. CMS Beliefs

We also appraised if pre-service teachers' self-efficacy beliefs for classroom management and perceptions of competence in specific CMS developed over time. Non-significant Shapiro-Wilk tests indicated normal distributions for these variables on both measurement points, which warranted the use of two separate paired-sample *t*-tests. For self-efficacy, we found that pre-service teacher, on average, held more positive beliefs towards their CMS in the post-test (M = 6.21, SD = 1.07) than in the pre-test (M = 5.53, SD = 1.05), $t(23) = -3.07$, $p = 0.005$, $d = 0.63$, 95% CI [−1.06, −0.18]. Moreover, pre-service teachers reported a higher overall CMS-competence level in the post-test (M = 3.98, SD = 0.30) compared to the pre-test (M = 3.60, SD = 0.35), $t(23) = -4.05$, $p \leq 0.001$, $d = 0.83$, 95% CI [−1.29, −0.35].

3.3. Degree of Alignment

Different instruments and data sources were used to capture pre-service teachers' over-time development of CMS behavior and CMS beliefs. Below, we will first reflect on the alignment of conclusions across the two time points and then discuss the degree of alignment in terms of over-time development.

With regard to pre-service teachers' self-reported use of CMS behaviors, we found that pre-service teachers, on average, thought they used the 17 CMS ranging from sometimes to regularly. This was the case for both the pre- and post-measurement, and, thus, does not indicate any over-time change. In terms of the type of CMS pre-service teachers' actual use (i.e., occurrence of CMS categorized in four CMS categories), in both sessions, we observed that deflection strategies were used most often, whereas reaction strategies were used least often. In addition, whereas anticipation, deflection, and reaction CMS were more often used in the second session, the use of ineffective CMS decreased over time. Regarding the adequacy of observed CMS behavior, the median of missed behaviors in the first session suggests that approximately one-sixth of all activated behaviors were not observed, whereas this was the case for approximately one in thirteen behaviors in the second session. Regarding the timeliness of observations, we found that pre-service teachers, on average, took approximately 9 and 7 s to respond to behaviors during the first and second sessions, respectively.

For the self-reported self-efficacy beliefs for CM, we see that pre-service teachers felt they were to some degree competent in CM, whereas, in the pre-measurement, they felt mostly competent. A similar pattern can be observed in pre-service teachers' perceived competence in the 17 specific CMS: at the pre-measurement, pre-service teachers, on average, felt somewhat competent in CMS, whereas they felt mostly competent at the time of the second measurement.

Thus, in terms of CMS behaviors, there is partial alignment. All observational measures of pre-service teachers' CMS behaviors (i.e., the type of CMS they used and the

adequacy and timeliness of their responses) suggest an over-time development of pre-service teachers' actual CMS usage. First of all, we found a significant increase in the use of anticipation, deflection, and reaction, but also ineffective CMS across sessions. In addition, the ratio of missed behaviors decreased significantly over time, indicating that pre-service teachers observed more (potentially) disruptive behaviors in the second session and, thus, were able to respond to and manage more behaviors as compared to the first session. Moreover, pre-service teachers responded noticeably faster to disruptive behaviors in Session 2 as compared to Session 1.

With regard to CMS beliefs, we can conclude that a positive development of CMS beliefs was captured by both self-reported measures: Pre-service teachers felt more efficacious and more competent in all 17 strategies after the second session as compared to the first session, signaling alignment between the two measures.

However, pre-service teachers' self-reported use of CMS behaviors and the actual use in the VR classroom diverge. More specifically, pre-service teachers' overall perceptions of the frequency in which they applied the 17 CMS when teaching in the (VR-)kindergarten classroom did not significantly change over time. This does not align with the objective instructor-based observations of pre-service teachers' CMS use, in which an increase in anticipation, deflection, and reaction strategies was observed. Here, it should be noted that the increase in the use of ineffective CMS is not mentioned as these CMS were not included in the self-report.

4. Discussion

In this article, we set out to evaluate how different assessment instruments may lead to aligning or diverging conclusions regarding VR-supported CMS development. We aimed to illustrate why triangulation of different assessment instruments can be conducive to adequately evaluating, understanding, and effectuating the added value of VR learning environments for training multifaceted skills such as CMS. Below, we will briefly discuss the variety of findings derived from multiple instruments across time points and reflect on the degree to which alignment of conclusions regarding CMS development is supported. Then, we will discuss why triangulation of instruments is needed if the goal is to adequately evaluate the potential of VR to support pre-service teachers' development of CMS behaviors and beliefs.

4.1. Assessing CMS Development Through Multiple Measurement Instruments

Six different indicators were used to capture pre-service teachers' CMS behaviors and CMS beliefs at two time points: VR-based observations of pre-service teachers' usage of the four CMS categories, the ratio of ignored or missed behaviors, the timelines of responses, and pre-service teachers' self-reported perceptions of CMS usage, self-efficacy beliefs, and perceived CMS-competence. From this variety of instruments, we can deduce that the direction and/or frequency of pre-service teachers' CMS behavior and CMS beliefs largely align across time points. For example, in both VR sessions, deflection CMS categories were used most often, whereas ineffective strategies were least often used in both VR sessions. In general, established theories on classroom management favor the usage of anticipation strategies, over-reactive and deflection strategies as anticipation CMS minimally disrupts a lesson's flow, which is known to benefit the learning climate (e.g., Borich, 2011; Kounin, 1970). However, for kindergarten teachers, it is imperative to give multiple stepwise directions when supporting young children in diverting or correcting misbehaviors (e.g., Ritz et al., 2014), which could explain the high occurrence of deflection strategies.

Regarding over-time development, the degree of alignment is less straightforward. For example, we found that pre-service teachers' overall perceptions of the frequency in which

they would apply the 17 CMS (that can be subsumed under anticipation, deflection, and reaction strategies) when teaching (real-life) kindergarten did not significantly change over time. One explanation could be that the self-reported use of anticipation strategies such as scanning increased, but development was counterbalanced by a reported decrease in some reactive strategies such as giving time-outs or removal. However, this was corroborated by the instructor-based observations of pre-service teachers' use of anticipation, deflection, and reactive CMS categories: we found a significant increase in the use of all three categories across sessions. Interesting to note is that the use of ineffective CMS did decrease over time.

One could question if an increase in reactive strategies is a positive development in terms of CMS behavior, as anticipation strategies are generally preferred. However, Bear (2015) found that authoritative reactive strategies, such as time-outs, can be more effective in decreasing future non-compliance than less intrusive reactive CMS, such as warnings. It could be that pre-service teachers perceived the behavioral scripts offered in Session 2 (in which more [complex] behaviors were activated) as more difficult and that the VR environment mostly prompted them to respond to behaviors as activated by the instructors—resulting in more positive perceptions towards the successfulness of more authoritative and reactive strategies. However, we did not gather data to systematically evaluate this hypothesis.

The development of pre-service teachers' CMS behaviors was also evidenced by a decrease in the ratio in which pre-service teachers ignored or missed behaviors and in their reaction times. These findings are of particular interest for teacher educators, as pre-service teachers, in general, fail or take longer to observe (potentially) disruptive behaviors compared to experienced and effective teachers (Van den Bogert et al., 2014), and early and accurate recognition facilitates timely implementation of CMS. To this end, being able to map (the development of) pre-service teachers' adequacy and timeliness in which they observe and respond to behaviors while training in a VR classroom has potential for teacher educators as it can provide valuable information on how to support pre-service teachers' development of CMS behaviors.

The self-report instruments allowed us to observe how pre-service teachers' self-efficacy beliefs towards more general classroom management approaches and their perceived competence in the 17 CMS changed over time. We found that at the post-test measurement, pre-service teachers, on average, reported more positive self-efficacy beliefs towards their classroom management skills and were more positive regarding their overall competence in the 17 CMS. This is a particularly encouraging finding as previous research documented a positive association between teachers' self-efficacy levels and their confidence and steadfastness towards challenging classroom management situations, and more efficacious teachers are more likely to believe that students' misbehaviors can improve (Dicke et al., 2014; Gordon, 2001). We acknowledge that, based on the data presented in the current article, we cannot make any definite conclusions regarding the effectiveness of VR for CMS learning as we were not able to compare self-reported and observational data of pre-service teachers in the VR group to data of their peers who did not participate in the VR-intervention. While interesting, such an effectiveness-oriented evaluation reaches beyond the goal of this manuscript as we wanted to illustrate how using multiple assessment instruments can reflect over-time development of multifaceted skills such as CMS and to understand how these instruments align. Therefore, we focused on self-reports and behavioral observations measured on two occasions, for which only the data of students allocated to the VR group were eligible.

4.2. Theoretical and Methodological Implications

Considering the high degree of alignment, one could question if multiple measurement instruments are needed or if fewer instruments would suffice in terms of content validity. Nevertheless, based on the findings presented in this article, we strongly advocate triangulation through a variety of instruments for three reasons. First, we believe that decisions regarding the usefulness and effectiveness of VR for professional skill training should not be based on half the story. Complex professional skills such as CMS entail many different aspects, and triangulating a variety of self-report measures with actual behavioral observations allowed us to gain insight into both the behavioral and beliefs components of CMS, helping us to better understand how pre-service teachers' perceptions of CMS and use are reflected in daily (virtual) teaching practices. To this end, acknowledging the multifaceted nature of complex skills prior to designing evaluation methods and the VR-learning environment itself seems a fruitful approach (Pantelidis, 2009). Moreover, we stipulate that if the goal is to understand how VR can contribute to the development of complex professional skills in higher education, research on the degree to which students respond to instructors' purposefully designed didactical learning opportunities is needed, as is a thoughtful reflection on the degree to which measurement instruments align the instructional and pedagogical theories underlying a learning design (Mouw & Fokkens-Bruinsma, 2022; Pantelidis, 2009; Radianti et al., 2020) We encourage future researchers to combine these insights with empirically grounded considerations of professional development (e.g., Sims et al., 2023) resulting in experimental settings in which we can determine effective characteristics of a VR-supported training setting. For example, in terms of the length and frequency of training sessions, the complexity of the behaviors, the type of activities (in addition to VR) that should be facilitated to promote reflection and development of specific competencies, etc. A solid evaluation of these factors helps in understanding how VR training environments can be organized and used for professional development.

Second, through triangulation, we can overcome prior issues in assessment as we no longer solely rely on either computer-generated data, knowledge tests, or self-report questionnaires, some of which have questionable validity (e.g., Carruth, 2017; Radianti et al., 2020; Renganayagalu et al., 2021). Triangulating self-report data with observational data can provide valuable insights with regard to different facets of CMS behavior and beliefs. For example, our coding schema allowed for capturing pre-service teachers' use of ineffective CMS (i.e., negative flow, negative with-itness, and ignoring or not observing behaviors) that are normally not included in self-questionnaires as these focus on self-efficacy beliefs towards CMS that are thought to be effective. Moreover, this study's easy-to-use measurement instruments resulted in medium to large effect sizes and explained considerable proportions of between-subjects variance, and as such, are promising to support researchers in their future decisions to employ both quantitative and qualitative evaluations to better understand the potential of VR for the development of professional skills (Radianti et al., 2020). Last, even though we could not evaluate the psychometric properties of all our instruments in our small-scale, innovation-based study, we believe that the instruments used in the present study have sufficient construct- and face validity as they mapped (the development of) distinct aspects of CMS (i.e., behaviors and beliefs). Using large-sample and fully crossed research designs, future studies could progress our work by applying a multitrait-multimethod approach to improve validity in a more structured manner.

Third, and perhaps most importantly, the triangulation of multiple instruments provides valuable insight into where pre-service teachers stand in the development of their CMS. Assessing different aspects of CMS before and after practicing these skills in a VR kindergarten could help instructors recognize individual and differential pathways in

which different aspects of pre-service teachers' CMS develop. From this point of view, triangulation is needed not only in terms of measurement instruments but also in terms of parallel analytical approaches. Including a person-oriented approach with latent profile analysis (see Hickendorff et al., 2018, for an introduction) may offer an additional perspective as it could help unravel individual characteristics serving as facilitating mechanisms for CMS development. Moreover, by evaluating over-time changes in profile membership, we can more adequately assess the differential potential of VR for specific students. For example, for pre-service teachers who struggle to manage kindergarteners and for natural kindergartner-teachers.

5. Conclusions

To conclude, this study's findings provide valuable theoretical and methodological insights into the added value of using multiple instruments to capture, assess, and evaluate VR-supported development of distinct aspects of pre-service teachers' kindergarten-specific CMS. In addition, we have illustrated how triangulation of different measurement instruments and assessment approaches can be used to overcome existing issues in skill assessment in simulation-based learning environments, allowing for an adequate measurement and understanding of the potential of VR for training CMS in future studies.

And even though our results can be interpreted as a 'positive development of pre-service teachers' kindergarten-specific CMS', which consequently may seem to hint towards a potential of VR kindergarten classrooms for developing CMS behaviors and beliefs, we do not want to give the false impression that a VR kindergarten classroom can be considered as a substitute for real-life classroom field experiences. What we do hope is that VR offers an additional playground in which pre-service teachers can fully grasp the reciprocal intricacies of specific CMS and kindergartners' behaviors, helping them to find their own CMS voice and discover what works for them. In the long term, this could lead to accelerated development of CMS—either throughout teacher education or as soon as they start as certified teachers in their own kindergarten classrooms. Understanding how VR-supported CMS development can be best assessed helps us to realize this goal as it can serve as a base for the provision of just-in-time tailored feedback supporting reflective practices that underly effective professional development.

Author Contributions: Conceptualization, J.M.M. and M.F.-B.; Methodology, J.M.M., M.F.-B. and A.S.; Formal analysis, J.M.M.; Investigation, J.M.M., M.F.-B. and A.S.; Resources, J.M.M. and A.S.; Data curation, J.M.M. and A.S.; Writing—original draft, J.M.M.; Writing—review & editing, J.M.M., M.F.-B. and A.S.; Supervision, J.M.M.; Project administration, J.M.M.; Funding acquisition, J.M.M. All authors have read and agreed to the published version of the manuscript.

Funding: The development of the virtual reality kindergarten classroom was realized by the Netherlands Initiative for Education Research 405.20865.262, and the Dutch Ministry for Education (OCW) through funding under the Open and Online Education Incentive Scheme OL22-02.

Institutional Review Board Statement: The Ethics Committee of Pedagogical and Educational Sciences at the University of Groningen PED-1920-S-0020, June 11, 2020.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The original contributions presented in the study are included in the article, further inquiries can be directed to the corresponding author.

Acknowledgments: We gratefully thank Thianda Spithoff, Ilona Stoffers, Geoffrey van der Steeg, and Steven Veldman for their assistance during the VR sessions and data collection.

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Appendix A

Table A1. Overview of Behavioral Scripts as Activated During VR session 1 and 2.

Script 1	Script 2	Script 3	Script 4
Face left	Face backwards	Stare at ceiling	Face right
Fiddle with clothes	Hand in pocket	Relaxing sit (Session 1: 1 kindergartner; Session 2: 2 kindergartners)	Look around (Session 1: 3 kindergartners; Session 2: 4 kindergartners)
Whisper left	Whisper right	Laughing	Teacher!
Scenario 5 (repeatedly jump off chair)	Scenario 11 (hit neighbor/laughing)	Scenario 12 (3 kindergartners talking/making noise)	Scenario 13 (fire truck passing by)
Whole-class unrest (Session 1: low; Session 2: mid-level)	Whole-class unrest (Session 1: low; Session 2: mid-level)	Whole-class unrest (Session1: low; Session 2: mid-level)	Whole-class unrest (Session1: low; Session 2: mid-level)
Beeping sounds	Clicking sounds (Session 1: 2 kindergartners; Session 2: 3 kindergartners)	Sing a song	Roar as a lion
Scenario 2 (whispering/talking)	Bathroom!	Scenario 7 (raise hand/talk out of turn)	Scenario 4 (stare at ceiling/turn backwards/look outside)
Stamping feet (4 kindergartners)	Stand up	Crawl (3 kindergartners)	Walk to table
Session 2: Whole-class unrest (mid- to high-level)	Session 2: Whole-class unrest (mid- to high-level)	Session 2: Whole-class unrest (mid- to high-level)	Session 2: Whole-class unrest (mid- to high-level)
Scenario 8 (crawlunder table/refuse to obey)	Scenario 14 (walk to table/pick up helmet)	Scenario 15 (2 kindergartners lose focus/stare at ceiling)	Scenario 10 (sit sideways/neighbors laugh)
Push/touch right neighbor (Session 2: 2 kindergartners)	Push/touch left neighbor	Stand up	Face backwards

Note. Underlined behaviors are only offered in Session 2.

Appendix B

Table A2. Correlations for all Measurements at Both Time Points.

	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.	14.	15.	16.	17.
1.	–																
2.	0.40 *	–															
3.	−0.07	−0.16	–														
4.	−0.22	0.08	0.15	–													
5.	0.20	0.03	−0.39	−0.14	–												
6.	−0.04	−0.06	−0.18	−0.39	0.17	–											
7.	−0.15	0.09	0.67	0.01	−0.09	0.04	–										
8.	0.42 *	0.23	0.04	−0.04	0.18	−0.41 *	0.24	–									
9.	0.13	0.07	−0.13	0.25	0.57 **	−0.003	0.06	0.05	–								
10.	−0.12	−0.06	0.15	−0.13	−0.08	0.12	−0.18	−0.34	−0.07	–							
11.	0.12	−0.10	−0.06	−0.51 *	0.17	0.84 **	0.13	−0.26	−0.06	−0.04	–						
12.	−0.39	−0.30	0.08	−0.21	0.04	0.24	−0.13	−0.53 **	−0.21	0.74 **	0.07	–					
13.	−0.02	0.01	−0.26	−0.34	0.49 *	0.46 *	0.03	−0.05	−0.07	−0.26	0.46 *	−0.07	–				
14.	−0.31	0.10	−0.12	0.05	0.50 *	0.26	0.81	−0.13	0.27	−0.12	0.14	0.001	0.53 *	–			
15.	0.15	0.33	−0.41	0.06	−0.08	0.19	−0.18	−0.13	−0.03	−0.29	0.30	−0.27	−0.12	−0.10	–		
16.	−0.05	0.44 *	−0.16	0.06	−0.03	−0.02	0.16	−0.09	0.05	−0.32	−0.03	−0.34	0.15	0.36	0.48 *	–	
17.	0.59 *	0.41 *	−0.14	0.33	0.19	0.07	0.15	0.43 *	0.002	−0.23	0.23	−0.30	0.02	−0.13	0.52 *	0.21	–
18.	0.06	0.35	−0.37	0.09	0.12	−0.14	−0.15	−0.03	0.07	−0.32	−0.15	−0.33	0.21	0.24	0.38	0.73 **	0.02

Note. 1 = CMS usage pre-test; 2 = CMS usage post-test; 3 = anticipation strategies Session 1; 4 = deflection strategies Session 1; 5 = reaction strategies Session 1; 6 = ineffective strategies Session 1; 7 = anticipation strategies Session 2; 8 = deflection strategies Session 2; 9 = reaction strategies Session 2; 10 = ineffective strategies Session 2; 11 = ratio ignored/missed behaviors Session 1; 12 = ratio ignored/missed behaviors Session 2; 13 = reaction time Session 1; 14 = reaction time Session 2; 15 = Self-efficacy CM pre-test; 16 = self-efficacy CM post-test; 17 = CMS competence pre-test; 18 = CMS competence post-test. * $p \leq 0.05$. ** $p \leq 0.01$.

References

Bauml, M. (2011). "We learned all about that in college": The role of teacher preparation in novice kindergarten/primary teachers' practice. *Journal of Early Childhood Teacher Education*, *32*(3), 225–239. [CrossRef]

Bear, G. (2015). Preventive and classroom-based strategies. In *Handbook of classroom management* (E. T. Emmer & E. J. Sabornie, Eds.). Routledge: pp. 15–39.

Billingsley, G., Smith, S., Smith, S., & Meritt, J. (2019). A systematic literature review of using immersive virtual reality technology in teacher education. *Journal of Interactive Learning Research*, *30*(1), 65–90.

Borich, G. (2011). *Effective teaching methods: Research-based practice*. Pearson Prentice Hall.

Brown, K., & Wynn, S. (2007). Teacher retention issues: How some principals are supporting and keeping new teachers. *Journal of School Leadership*, *17*, 664–698. [CrossRef]

Burdea, G., & Coiffet, P. (2003). *Virtual Reality technology*, 2nd ed. Wiley.

Carruth, D. (2017). Virtual reality for education and workforce training *[Paper presentation]*. International Conference on Emerging eLearning Technologies and Applications 15th Annual Meeting, Stary Smokovec, Slovakia, October 26–27. [CrossRef]

Carter, K., & Doyle, W. (2006). Classroom management in early childhood and elementary classrooms. In *Handbook of classroom management, research, practice, and contemporary issues* (C. M. Evertson & C. S. Weinstein, Eds.). Routledge: pp. 373–406. [CrossRef]

Chernikova, O., Heitzmann, N., Stadler, M., Holzberger, D., Seidel, T., & Fischer, F. (2020). Simulation-Based learning in higher education: A meta-analysis. *Review of Educational Research*, *7*, 499–541. [CrossRef]

Christofferson, M., & Sullivan, A. (2015). Preservice teachers' classroom management training: A survey of self-reported training experiences, content coverage, and preparedness. *Psychology in the Schools*, *52*(3), 248–264. [CrossRef]

Dicke, T., Parker, P. D., Marsh, H. W., Kunter, M., Schmeck, A., & Leutner, D. ((2014). Self-efficacy in classroom management, classroom disturbances, and emotional exhaustion: A moderated mediation analysis of teacher candidates. *Journal of Educational Psychology*, *106*(2), 569–583. [CrossRef]

Dobbs-Oates, J., Kaderavek, J. N., Guo, Y., & Justice, L. M. (2011). Effective behavior management in preschool classrooms and children's task orientation: Enhancing emergent literacy and language development. *Early Childhood Research Quarterly*, *26*, 420–429. [CrossRef]

Eisenman, G., Edwards, S., & Cushman, C. (2015). Bringing reality to classroom management in teacher education. *The Professional Educator*, *39*(1), 1–12.

Emmer, E. T., & Stough, L. M. (2001). Classroom management: A critical part of educational psychology, with implications for teacher education. *Educational Psychologist*, *36*(2), 103–112. [CrossRef]

Flower, A., McKenns, J. W., & Haring, C. D. (2017). Behavior and classroom management: Are teacher preparation programs really preparing our teachers? *Preventing School Failure: Alternative Education for Children and Youth*, *61*(2), 163–169. [CrossRef]

Gordon, L. M. (2001). High teacher efficacy as a marker of teacher effectiveness in the domain of classroom management *[Paper presentation]*. Annual Meeting of the California Council on Teacher Education, San Diego, CA, USA, October 25–27.

Hickendorff, M., Edelsbrunner, P., McMullen, J., Schneider, M., & Trezise, K. (2018). Informative tools for characterizing individual differences in learning: Latent Class, Latent Profile, and Latent Transition Analysis. *Learning and Individual Differences*, *66*, 4–15. [CrossRef]

Huang, Y., Richter, E., Kleickmann, T., Wiepke, A., & Richter, D. (2021). Classroom complexity affects student teachers' behavior in a VR classroom. *Computers & Education*, *163*, 104100. [CrossRef]

Jensen, L., & Konradsen, F. (2018). A review of the use of virtual reality head-mounted displays in education and training. *Education and Information Technologies*, *23*, 1515–1529. [CrossRef]

Jones, V. (2006). How do teachers learn to be effective classroom managers? In *Handbook of classroom management, research, practice, and contemporary issues* (C. M. Evertson & C. S. Weinstein, Eds.). Routledge: pp. 887–907.

Korpershoek, H., Harms, T., De Boer, H., Van Kuijk, M., & Doolaard, S. (2016). Meta-analysis of the effects of classroom management strategies and classroom management programs on students' academic, behavioral, emotional, and motivational outcomes. *Review of Educational Research*, *86*, 643–680. [CrossRef]

Korpershoek, H., Mouw, J. M., & De Boer, H. (2022). International research on the foci and effectiveness of classroom management programs and strategies. In *Handbook of classroom management*, 3rd ed. (E. J. Sabornie & D. L. Espelage, Eds.). Routledge: pp. 350–372.

Kounin, J. S. (1970). *Discipline and group management in classrooms*. Holt, Rinehart, & Winston.

Lee, E., & Wong, K. (2008). A review of using virtual reality for learning. In *Transactions on edutainment I* (Z. Pan, A. D. Cheok & W. Müller, Eds.). Springer: pp. 231–241.

Lugrin, J.-L., Latoschik, M., Habel, M., Roth, D., Seufert, C., & Grafe, S. (2016). Breaking bad behaviors: A new tool for learning classroom management using Virtual Reality. *Frontiers in ICT*, *3*, 1–21. [CrossRef]

McGarr, O. (2021). The use of virtual simulations in teacher education to develop pre-service teachers' behaviour and classroom management skills: Implications for reflective practice. *Journal of Education for Teaching*, *47*(2), 274–286. [CrossRef]

McGrath, J. L., Taekman, J. M., Dev, P., Danforth, D. R., Mohan, D., Kman, N., Crichlow, A., & Bond, W. F. (2018). Using virtual reality simulation environments to assess competence for emergency medicine learners. *Academic Emergency Medicine*, *25*, 186–195. [CrossRef]

McHugh, M. L. (2012). Interrater reliability: The kappa statistic. *Biochemia Medica*, *22*, 276–282. [CrossRef]

Mouw, J. M., & Fokkens-Bruinsma, M. (2022). When technology meets educational sciences: Combining virtual reality and microteaching to train pre-service teachers' kindergarten classroom management strategies. In *Proceedings of the 8th International Conference on Higher Education Advances* (J. Domenech, Ed.). Universitat Politècnica de València: pp. 1043–1050. [CrossRef]

Mouw, J. M., Fokkens-Bruinsma, M., & Verheij, G.-J. (2020). Using Virtual Reality to promote pre-service teachers' classroom management skills and teacher resilience: A qualitative evaluation. In *Proceedings of the 6th International Conference on Higher Education Advances* (J. Domenech, P. Merello, E. de la Poza & R. Peña-Ortiz, Eds.). Universitat Politècnica de València: pp. 325–332. [CrossRef]

Pantelidis, V. S. (2009). Reasons to use virtual reality in education and training courses and a model to determine when to use virtual reality. *Themes in Science and Technology Education*, *2*(1–2), 59–710.

Poulou, M. S., Reddy, L. A., & Dudek, C. M. (2019). Relation of teacher self-efficacy and classroom practices: A preliminary investigation. *School Psychology International*, *40*(1), 25–48. [CrossRef]

Putman, S. M. (2009). Grappling with classroom management: The orientations of preservice teachers and impact of student teaching. *The Teacher Educator*, *44*, 232–247. [CrossRef]

Radianti, J., Majchrzak, T., Fromm, J., & Wohlgenannt, I. (2020). A systematic review of immersive virtual reality applications for higher education: Design elements, lessons learned, and research agenda. *Computers & Education*, *147*, 103778. [CrossRef]

Renganayagalu, S. K., Mallam, S. C., & Nazir, S. (2021). Effectiveness of VR head mounted displays in professional training: A systematic review. *Technology, Knowledge and Learning*, *26*, 999–1041. [CrossRef]

Ritz, M., Noltemeyer, A., Davis, D., & Green, J. (2014). Behavior management in preschool classrooms: Insights revealed through systematic observation and interview. *Psychology in the Schools*, *51*, 181–197. [CrossRef]

Simonsen, B., Fairbanks, S., Briesch, A., Myers, D., & Sugai, G. (2008). Evidence-based practices in classroom management: Considerations for research to practice. *Education and Treatment of Children*, *31*(3), 351–380. [CrossRef]

Sims, S., Fletcher-Wood, H., O'Mara-Eves, A., Cottingham, S., Stansfield, C., Goodrich, J., Van Herwegen, J., & Anders, J. (2023). Effective teacher professional development: New theory and a meta-analytic test. *Review of Educational Research*. [CrossRef]

Theelen, H., Van den Beemt, A., & Den Brok, P. (2019). Classroom simulations in teacher education to support preservice teachers' interpersonal competence: A systematic literature review. *Computers & Education*, *129*, 14–26. [CrossRef]

Tschannen-Moran, M., & Woolfolk-Hoy, A. (2001). Teacher efficacy: Capturing an elusive construct. *Teaching and Teacher Education*, *17*(7), 783–805. [CrossRef]

Van den Bogert, N., Van Bruggen, J., Kostons, D., & Jochems, W. (2014). First steps into understanding teachers' visual perception of classroom events. *Teaching and Teacher Education*, *37*, 208–216. [CrossRef]

Van der Want, A. C., & Visscher, A. J. (2024). Virtual Reality in preservice teacher education: Core features, advantages and effects. *Educational Sciences*, *14*, 635. [CrossRef]

Van Tartwijk, J., Veldman, I., & Verloop, N. (2011). Classroom management in a Dutch teacher education program: A realistic approach. *Teaching Education*, *22*, 169–184. [CrossRef]

Vatcheva, K. P., Lee, M. J., McCormick, J. B., & Rahbar, M. H. (2016). Multicollinearity in regression analyses conducted in epidemiologic studies. *Epidemiology (Sunnyvale)*, *6*(2), 227. [CrossRef]

Zeichner, K. M., & Tabachnick, B. R. (1981). Are the effects of university teacher education "washed out" by school experience? *Journal of Teacher Education*, *3*, 7–11. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.