Evaluating the validity of SimTeach lessons as an alternative to face-to-face classroom observations

Sabrina Syed
Charles Sturt University, Port Macquarie, Australia; The University of Newcastle, Callaghan, Australia

Claire Bates, Elena Prieto-Rodriguez, Susan Ledger
The University of Newcastle, Callaghan, Australia

Class observations are a widely used form of professional development. They are also used for measuring teaching quality of both in-service and pre-service teachers. While they have demonstrated benefits when utilised as a form of professional development, if used for evaluative purposes, they can be overwhelming for teachers, students and observers alike. This study aims to explore whether observing simulated lessons is a valid alternative to traditional physical classroom observations. To do so, we analysed 20 recorded mathematics lessons taught using SimTeach, a virtual simulation platform. The observations employed the quality teaching model as a framework for measuring teaching quality. The analysis provided us with a measure of participants' teaching quality and revealed no significant differences between the two forms of observation. Though simulated lessons have proven effective in assessing teaching quality, direct observation in a physical classroom remains essential for a comprehensive understanding of student learning. This is because student engagement in a simulated environment is not entirely authentic, as responses are provided by professional actors and class sizes limited to five. While these responses may resemble those of real students, they do not fully capture the genuine interactions, emotions and learning dynamics present in an actual classroom setting.

Implications for practice or policy

- Live simulation experiences enhance teaching effectiveness and serve as a valuable tool for teachers' professional development.
- Simulated platforms present a promising alternative to traditional classroom observations for measuring teaching quality.
- Initial teacher educators should consider simulated lesson observations as an alternative to providing feedback to students on teaching practice.

Keywords: virtual teaching simulations, class observation, quantitative.

Introduction

There is widespread agreement that class observations are one of the most common and reliable methods for evaluating teaching quality (Cohen & Goldhaber, 2016; Hsiao et al., 2022). Hsiao et al. (2022) further acknowledges the value of class observation by stating that it provides actionable feedback and offers concrete insights into teacher performance and student engagement, which is crucial for both professional development and educational research. Observations have the potential to yield information that can challenge theorists to be more comprehensive or practitioners to improve or modify practice.

Classroom observations can be conducted face-to-face, remotely, or via video recordings. Face-to-face or physical class observations provide observers with more control over what to observe and allow clearer views of teacher-student and student-student interactions. However, such observations can be intrusive, potentially affecting the authenticity of the lesson being observed (Liang, 2015). One factor that may compromise the authenticity of face-to-face class observations is that teachers and students are often aware of the observations, which can unintentionally lead to unnatural behaviours or inaccuracies in the lesson (Waxman et al., 2004). For example, students tend to behave in a more reserved and unusual manner when aware they are being observed (Howard, 2010). Students may attempt to meet the perceived expectations of the observer, with their behaviour influenced by how they interpret the observer's role during the observation. Additionally, students who are not the focus of the observation may become distracted by the observer, sometimes attempting to show off or gain attention, further skewing the authenticity of the learning environment (Whitcomb & Merrell, 2013).

In contrast, remote or recorded observations eliminate the presence of an in-person observer in the classroom (Heafner et al., 2011), which can alleviate some of the stress associated with having an additional person physically present. However, the use of recording devices may introduce a different type of pressure, as the constant awareness of being recorded can create a heightened sense of surveillance and scrutiny (Richards, 2003). Teachers may find video recording intrusive or uncomfortable (Blazar et al., 2018). *Reactivity* is the phenomenon in which individuals being observed change their behaviour simply because they are being observed (Merrett, 2006). Reactivity is applicable in both face-to-face and recorded observations. These behaviour changes can bring errors to experimental conclusions (Miltenberger, 2016), similar to face-to-face observations. Additionally, remote or recorded observations pose challenges in perceiving facial expressions and distinguishing student voices during group discussions (Heafner et al., 2011).

Though there are contextual differences in these two modes of class observations, Heafner et al. (2011) found no significant difference in professional improvement outcomes between the two methods for graduate interns while using the *observation feedback form* (Heafner et al., 2011) and the *interstate teacher assessment and support consortium standards* (Council of Chief State School Officers, n.d.). This suggests that while physical and recorded observations are not equivalent, they are comparable, each with its own set of advantages and limitations. As such, Heafner et al. (2011) concluded that both observation methods could be used, as neither proved to be inherently superior for supporting professional growth.

There is now a potential alternative method of observing teaching practice that has emerged in recent times due to advances in technology: a simulation technology known as SimTeach (Mursion ™). SimTeach is a live simulation experience where student teachers can practice instructional strategies with avatars operated by a real human. We measured whether recorded SimTeach lessons could be a valid alternative to physical or recorded observations by utilising the New South Wales (NSW) quality teaching model (QTM) as an evaluation tool. The QTM is a framework used to evaluate and improve teaching practice, which focuses on three key dimensions: (a) intellectual quality (deep understanding of concepts), (b) quality learning environment (productive classroom focused on learning) and (c) significance (making learning meaningful for students) (Gore et al., 2021; Newmann, 1996). Together, these dimensions provide a comprehensive view of effective teaching practices.

If the goal is to understand and explore teaching strategies, SimTeach lessons may offer a solution to addressing issues related to the aforementioned classroom or remote observations. Simulations can minimise disruptions to regular classroom activities while still providing comparable insights into a teacher's instructional effectiveness. In the following section, we will further explore SimTeach lessons, microteaching and the QTM, as this study specifically observed SimTeach microteaching lessons and evaluated teacher performance using the QTM.

Simulation technologies in teacher education

Simulations are technologies that allows teachers to practice pedagogical skills in a low-stakes environment (Lew et al., 2021). The use of simulations in teacher education is relatively new but gaining popularity as a tool to enhance pre-service teacher education. Pre-service teachers often value learning from this type of virtual experience (Lew et al., 2021) because it provides a realistic teaching environment in which teachers' individual skills and teaching styles can be demonstrated (Berg et al., 2023). Simulation technologies provide the opportunity for additional preparation for pre-service teachers' practicums as they provide the ability to make mistakes and reflect without any negative impact on students (Fischetti et al., 2022). They can also be used as a tool to increase confidence, self-efficacy, classroom management skills and communication (Ade-Ojo et al., 2022). They allow teachers to improve their practice through an intensified reflective process (Berg et al., 2023). The main advantage of using simulations is to help teachers build and strengthen the skills needed for effective teaching (Fischetti et al., 2022).

While simulation technologies offer numerous advantages, such as the ability to aid in learning the basic aspects of teaching, they can be met with reluctance in initial teacher education due to potential technical challenges that can arise during implementation (Dittrich et al., 2022; Lindberg & Jönsson, 2023). These difficulties include system malfunctions, software compatibility issues, or the need for specialised training, all of which can impact the smooth integration of simulations into teaching and learning environments. Limitations further include a reported lack of authenticity, difficulties in engaging with the virtual avatars and difficulties in preparing for, conducting and reflecting on the experience (Ledger, 2019). Another potential limitation stems from the mode of delivery, as simulations are conducted either online or in a designated room that replicates traditional face-to-face teaching. This shift in format may reduce the direct, in-person interaction between instructors and students, which some educators might find less effective for developing certain practical teaching skills (Taylor, 2024). This challenge can be more pronounced in subjects like mathematics, which often rely heavily on hands-on activities and dynamic interactions between teachers and students. The need for immediate feedback, problem-solving collaboration and physical engagement with materials is harder to replicate in a virtual space, making it more difficult to foster the same depth of understanding and connection as in a traditional classroom setting (Taylor, 2024).

Simulated environments have benefits comparable to online and remote teaching contexts as well as similar technological, pedagogical and social challenges (Gillett-Swan, 2017). Technological challenges mainly relate to the unreliability of internet connections and a lack of necessary electronic devices. Pedagogical challenges are associated with digital skills, effective online resources, interactivity, motivation and lack of social and cognitive presence. Social challenges involve issues with group work and the lack of human interaction between teachers and students (Ferri et al., 2020). These challenges are also applicable in mathematics teaching. Amedu and Hollebrands (2022) highlighted significant challenges faced by mathematics teachers in teaching remotely, including the availability of online resources, difficulties with adapting pedagogy to an online format and a lack of interaction and feedback from students. Similarly, Yohannes et al. (2021) found that 70% of mathematics teachers identified geometry and trigonometry as the most challenging topics to teach online. These challenges are relevant to simulated teaching environments, including access to devices, internet connectivity issues, student engagement difficulties and concerns over the authenticity of student work (Ziebell et al., 2020). Conversely, Ziebell et al. (2020) found that online teaching, which can be extended to simulated teaching environments, offered benefits like fewer disruptions and increased creativity in lesson planning and teaching methods.

Most simulated classes replicate microteaching scenarios. Microteaching was first developed in 1963 at Stanford University and is defined as a teaching situation which is scaled down in terms of time and number of students (Allen & Clark, 1967). Microteaching scenarios can be four to 24 minutes long with a small number of students, approximately four to five, to allow a teacher to focus on selected aspects of teaching (Cooper & Allen, 1970). Similarly, for simulated teaching environments, teacher education students typically design a microteaching lesson, which they then practice in a simulated environment. This setup allows them to focus on specific teaching strategies, refine their instructional methods and receive feedback in a controlled and focused context before engaging in full classroom teaching.

Mukuka and Alex's (2024) review of microteaching in mathematics teacher education found that microteaching is an effective approach for improving math teaching skills among pre-service teachers.

Microteaching proved to be an effective diagnostic tool for identifying the specific needs of pre-service teachers and a preparatory tool for real-life placements. Performance in a microteaching situation has been found to accurately predict classroom performance (Cooper & Allen, 1970). However, the overcrowded teacher education curriculum has led many initial teacher education programmes to discontinue the use of microteaching. Therefore, engaging with simulation technology can became an alternative option for pre-service teachers in teacher education programmes (Fischetti et al., 2022). Unlike face-to-face microteaching, simulation teaching happens in a computerised environment involving a human interactor working behind the scenes to control the student avatars (Ledger, 2019). There are a range of simulations and mixed-reality learning environments available for educational purposes. They can be categorised as single-user programs, multi-user virtual environments and mixed-reality virtual simulations. Simulations that use avatars within a virtual learning environment allow teachers to feel that they are present within the virtual environment as they teach. Berg et al. (2023) found that microteaching in a simulated environment offered several benefits, including enhanced self-efficacy among pre-service teachers, its use as a valuable diagnostic tool for academics and better preparation of graduates for the workforce. However, challenges were primarily associated with feelings of being overwhelmed by the technology or perceptions that the lessons lacked realism (Ledger, 2019). Fischetti et al. (2022) advocated for the importance of microteaching as a strategy to practice before placement and highlighted the need for more widespread use of simulation in initial teacher education programmes.

The SimTeach platform is a single-user virtual environment simulation technology which provides an avatar-based virtual teaching and learning environment for practicing teaching methods, skills and language. It utilises a “human-in-the-loop” method, in which digital student avatars are controlled in real time by a human working behind the scenes (Lindberg & Jönsson, 2023). Compared to other simulation-based tools, SimTeach offers a structured environment for practicing classroom management and instructional strategies (Rosati-Peterson et al., 2021), yet its role can be situated within a wider ecosystem of digital interventions that support pedagogical innovation. It can be positioned within broader debates on the role of simulation and immersive technologies in teacher education and digital pedagogy. While traditional approaches rely on in-person observation and mentoring, simulation-based tools offer opportunities for safe, repeatable and reflective practice. Compared to other educational technologies, such as virtual classroom platforms (Turoff, 1995) or role-playing software (Brucklacher & Gimbert, 1999), SimTeach focuses specifically on observation and narrative-based reflection, enabling teacher candidates to critically engage with classroom scenarios and their own professional development. Although it does not currently incorporate AI-driven feedback systems, SimTeach contributes to the growing landscape of immersive learning tools by fostering reflective practice, supporting professional growth and providing a structured yet flexible environment for pre-service teachers. This positions our study as part of a wider conversation on how technology can enhance teacher preparation, complementing both traditional pedagogical approaches and other emerging educational technology solutions. Therefore, this study positions SimTeach not merely as a substitute for in-person observation but as part of a broader discussion on leveraging technology to enhance teacher preparation, professional development and reflective practice in higher education.

Quality teaching model

In order to enable comparisons between face-to-face and simulated environments, this study employed the QTM. We used the QTM model to analyse the teaching strategies of our participants and compare their practices across face-to-face and SimTeach formats. Our goal was to evaluate the simulation platform, specifically examining how similar and reliable it is in comparison to face-to-face lesson observations. The QTM has been used extensively for course planning, collegial collaboration and improving the student experience (Patfield et al., 2025). It is grounded in decades of research and has been used to assess key aspects of teaching across subjects from elementary to tertiary levels (Gore et al., 2021).

The QTM is outlined in two key documents: The *Quality Teaching Classroom Practice* (QT Academy, 2020) and the *Quality Teaching Assessment Practice Guide* (NSW Department of Education and Training, 2003; Quality Teaching (QT) Academy, 2020). This model was particularly suited to this study as it applies across various subjects and grade levels and is endorsed by the NSW Department of Education (Gore et al., 2017; Gore et al., 2021). The QTM is organised into three dimensions, each containing six elements. These elements are assessed using a 1-to-5 coding scheme, which serves as an observation tool for analysing pedagogical skill in both statistical and non-statistical terms (Harper, 2023).

The intellectual quality dimension emphasises active and extended engagement with the material. This involves sustaining focus on key concepts throughout the lesson (deep knowledge), student comprehension of the material (deep knowledge), students' engagement with critical thinking (higher-order thinking), in-depth discussions (substantive communication), the development of students' language (metalanguage) and encouraging students to question knowledge (problematic knowledge). Intellectual quality blends content and process-focused approaches, suggesting that both can be of high or low quality. The quality learning environment dimension focuses on creating classrooms where students are highly engaged (engagement), feel supported by both their teachers and peers (social support) and are self-directed (student direction) and self-regulated (student self-regulation). In these classrooms, students have internalised the rules and expectations, which is seen as a sign of high quality. Additionally, the model values classrooms with high expectations, where students are clearly told what is expected of them and where the standards for quality work are clearly explained and reinforced (explicit quality criteria). The dimension of significance connects both the social and psychological aspects of learning. On one side, students benefit individually from lessons that build on their prior knowledge (background knowledge), make connections across subject areas (knowledge integration) and relate classroom learning to real-world applications (connectedness). On the other side, students' social and cultural backgrounds are seen as valuable assets for learning. This includes recognising knowledge from non-dominant social groups (cultural knowledge), using narrative-based forms of knowledge (narrative) and ensuring that students from all backgrounds are engaged in classroom discussions (inclusivity). Both psychological and sociological factors are essential in making learning meaningful and significant for students (Ladwig, 2009).

Methodology

Our research goal was to explore whether recorded SimTeach lessons could serve as a valid alternative to traditional face-to-face class observations. There were no financial, commercial, or personal relationships with the company providing SimTeach that could have influenced the work reported in this study. We used quantitative data in the form of codes obtained during observations using the framework presented in the *Quality Teaching Classroom Practice Guide* (QT Academy, 2020). We coded SimTeach lessons and compared these codes with data from face-to-face observations conducted by the Teachers and Teaching Research Centre using the same framework. It is important to note that the QTM was designed for real classroom observations, which presented certain challenges when applying it in a simulated environment like SimTeach. In SimTeach, there are no real students; instead, avatars are controlled by an actor behind the scenes (University of Newcastle, 2023). According to the QTM, the quality learning environment dimension should be coded by observing student behaviours, particularly when coding engagement, social support and students' self-regulation (QT Academy, 2020). While the avatars behave in a way that closely mimics real students, their actions are ultimately determined by the human operator. Although there is a way to preset student behaviour levels to high, moderate, or low engagement, the authenticity of these interactions remains influenced by the individual controlling the avatars. Therefore, even if we decide to retain the engagement, social support and students' self-regulation codes in the paper, readers should be aware of the limitations this may pose when making comparisons. While student behaviour in a simulated environment may not perfectly mirror real classroom dynamics, such observations can still serve as a valuable tool for supporting teachers' professional development.

We observed and analysed 20 recorded SimTeach mathematics lessons collected from 17 participants in 2023 and 2024. The sample included five males (29%) and 12 females (71%) (representative of the gender balance in the programme) with all students enrolled in a postgraduate mathematics teacher education programme. The SimTeach lessons featured a 10-minute Year 8 geometry lesson on angles. Observers made notes by hand in relation to the teachers' instruction and lesson activities. Following the quality teaching rounds format (QT Academy, 2020), observers first watched the lessons and took handwritten notes, then individually coded each lesson immediately after viewing the recording before discussing and reaching a consensus on the final codes. The recordings were coded by three observers over six weeks from July to August 2024. Eight lessons were coded by all three observers, while 12 were coded by two. This step generated quantitative data in the form of a code from one to five awarded for each of the 18 pedagogical elements, with a code of one indicating that the element was not present in the lesson and a code of five indicating that the element was incorporated to a high standard throughout the lesson. It is important to note that student responses and facial expressions play a crucial role in observations. However, interpreting these cues was challenging, as the virtual characters are basic 2D animations with limited expressive detail. While understanding their expressions was not always be straightforward, it was still possible with careful observation.

We compared our SimTeach observation codes with those reported in three different datasets. It is important to note that the lessons included in this study were selected based on availability and feasibility, providing sufficient data to identify meaningful patterns in the comparative analysis of SimTeach and in-person observations.

1. Harper's (2023) face-to-face observations. Harper (2023) conducted a study examining mathematics lessons delivered in person by a single teacher for Year 10 and Year 11 students. The Year 10 class comprised 23 students, while the Year 11 class had 18 students. These observations provided valuable insights into traditional classroom instruction, student engagement and teaching strategies.
2. Findings from our 2023 face-to-face lesson observations. We observed five in-service, out-of-field teachers participating in a postgraduate mathematics retraining programme, all of whom were in the final stages of their degree. Their SimTeach lessons were also part of our coding.
3. SimTeach codes from a large-scale dataset. We examined SimTeach observation codes from approximately 280 recorded lessons delivered by in-service primary and secondary mathematics teachers in Australia. These observations were conducted as part of the Building Capacity for Quality Teaching in Australian Schools project (Gore et al., 2023).

All face-to-face lessons were analysed following the same coding procedures and employed the QTM. We compared the means of 18 elements across both modes. The study adhered to the university's ethics protocol, obtaining consent from participants for all observations. We utilised experienced coders who had applied the model across two earlier randomised controlled trials. The inter-rater reliability from the first of these trials is reported in Gore et al. (2017)'s publication. We also calculated the interclass correlation coefficient to assess inter-rater reliability for the current study, which resulted in a score of 0.67 (67%), indicating moderate agreement between coders. While this reflects reasonable consistency, some variation in judgments remained. Our primary aim was not to match individual ratings exactly but to reach a mutually agreed-upon coding. In several instances, we re-watched the videos together until both observers reached consensus. For the study, only these agreed-upon codes were used, rather than the individual coders' initial ratings. It is important to note that QTM codes serve primarily as a framework for facilitating professional dialogue and building consensus, rather than as a rigid rating system (Patfield et al., 2025; QT Academy, 2020). As Patfield et al. (2025) note, teachers using the model are encouraged to move beyond simply assigning numerical codes or rushing to agreement. Instead, the value lies in the depth of discussion itself, including challenging assumptions and carefully considering divergent viewpoints. Given this focus on professional dialogue over numerical scoring, traditional inter-rater reliability measures may not be particularly meaningful in this context. We also acknowledge that simulation fidelity may differ from live classroom dynamics, potentially influencing observed interactions and teaching behaviours. Additionally, this study focused exclusively on mathematics lessons, which may limit the generalisability of the findings to other subjects. Future research could extend this approach across a broader range of disciplines to strengthen external validity. Despite these limitations, the methodology provides a robust framework for comparing simulated and in-person observations and contributes to understanding the potential of simulation tools in teacher education.

Results and discussion

To analyse mathematics teaching strategies with the intention to understand the reliability of the SimTeach platform for class observation, we utilised the three domains of the QTM, each consisting of six dimensions, resulting in a total of scores from 18 dimensions for comparison. We systematically compared the scores from SimTeach observations with those from face-to-face lessons.

Intellectual quality

We gained valuable insights through our observations of SimTeach into our participants' teaching practices within the intellectual quality domain, highlighting both their strengths and areas for improvement. Participants excelled in dimensions such as deep knowledge (4.25 out of 5) and substantive communication (4.15 out of 5), indicating a strong grasp of subject content and the ability to engage in meaningful, substantive discussions. However, the results for problematic knowledge (1.35 out of 5) suggests a challenge in addressing and embracing the complexities or uncertainties inherent in the subject matter, an area crucial for fostering critical thinking and deeper understanding among students.

The comparison of our results with Harper (2023) is presented in Figure 1. In both cases, findings in the intellectual quality domain provide valuable insights into the strengths and weaknesses of participants in mathematics teaching. Harper's (2023) study observed face-to-face mathematics classes taught by one teacher for Year 10 and 11 students. The Year 10 class was comprised of 23 students, while the Year 11 class included 18 students. Although the exact duration of the classes was not specified, Harper (2023) observed seven Year 10 classes and 22 Year 11 classes, totalling 48.75 hours of observation. It is important to note that this total encompasses both mathematics and drama classes.

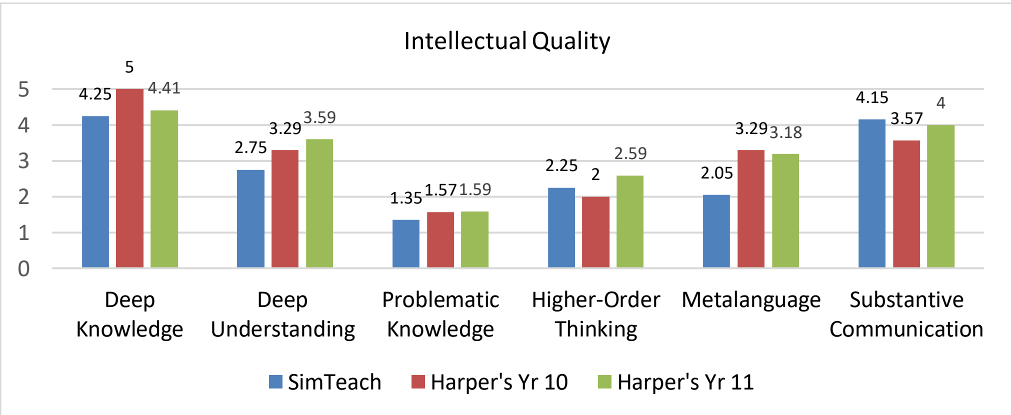

Figure 1. Comparison of the average codes awarded to the SimTeach lessons and Harper's (2023) Year 10 and 11 lessons for each element of the intellectual quality dimension of the QTM.

In Harper's (2023) study, the instruction was not implicit; explicit instruction and tightly bound content (visible pedagogy) led to higher intellectual quality but lower significance codes in mathematics. The findings revealed that codes for deep knowledge (5 in Year 10 and 4.41 in Year 11) and substantive communication (3.57 in Year 10 and 4 in Year 11) were high in mathematics, similar to our results. However, both studies identified a lack of engagement with problematic knowledge. In Harper's (2023) study, mathematics teachers received an average of 1.57 in Year 10 and 1.59 in Year 11 for problematic knowledge. In our study, the average code was 1.35 for problematic knowledge. This sheds light on a key contrast: mathematics adheres to structured, discipline-specific content. Mathematics teachers found it challenging to encourage higher-order thinking in both face-to-face and SimTeach lessons. However, the codes for substantive communication and deep understanding were consistently above 3.5. Although some SimTeach codes were lower than the face-to-face observations, the differences were not significant. We acknowledge that the SimTeach observations occurred in a microteaching context, while the face-to-face observations took place in full-length classes, which may account for the differences. For example, metalanguage was coded slightly above 2 in the SimTeach lesson observations, whereas Harper's (2023) study recorded a code of over 3.

The differences may reflect the contextual constraints of the SimTeach microteaching sessions, which were shorter in duration, involved fewer students and lacked the dynamic interactions of full-length classes. Certain aspects of the QTM dimensions in teaching practice may be particularly sensitive to the simulation context, such as problematic knowledge and metalanguage. One reason is that simulated lessons are short and teachers may not plan activities that encourage students to question knowledge. Additionally, teachers are aware that the participants are actors rather than actual students, which can influence their approach. Similarly, the brief duration affects metalanguage: to achieve a higher code in this dimension, a teacher would need to conduct a small session explicitly focused on explaining language concepts. In practice, teachers tend to prioritise mathematics content rather than incorporating a short language-focused segment in SimTeach. Cognitive load (Kirschner et al., 2018), the novelty of the virtual environment and reduced social presence could have further contributed to these variations, influencing teachers' confidence and willingness to engage with more complex or abstract content. Additionally, the differences highlight the potential limitations of using simulation as a sole measure of teaching quality, particularly for higher-order thinking skills that rely on real-time negotiation and student-teacher interaction. Overall, these findings indicate that while SimTeach is effective in capturing core elements of teaching quality, contextual factors inherent to the simulated environment may influence performance in specific dimensions. SimTeach therefore is best positioned as a complementary tool to face-to-face observation, particularly when assessing dimensions that require dynamic classroom interaction, problem-solving and social support.

Quality learning environment

The comparison of our results with Harper's (2023) findings also reveals significant insights into the dynamics of the quality learning environment across different educational settings. As mentioned earlier, this is the dimension where we needed to ensure student engagement by observing their expressions and responses. Since it was a simulated environment and the participants were not actual students, we must be cautious when interpreting the findings from this section. However, we decided to include them in the paper because the simulation closely mirrored the behaviour of real students.

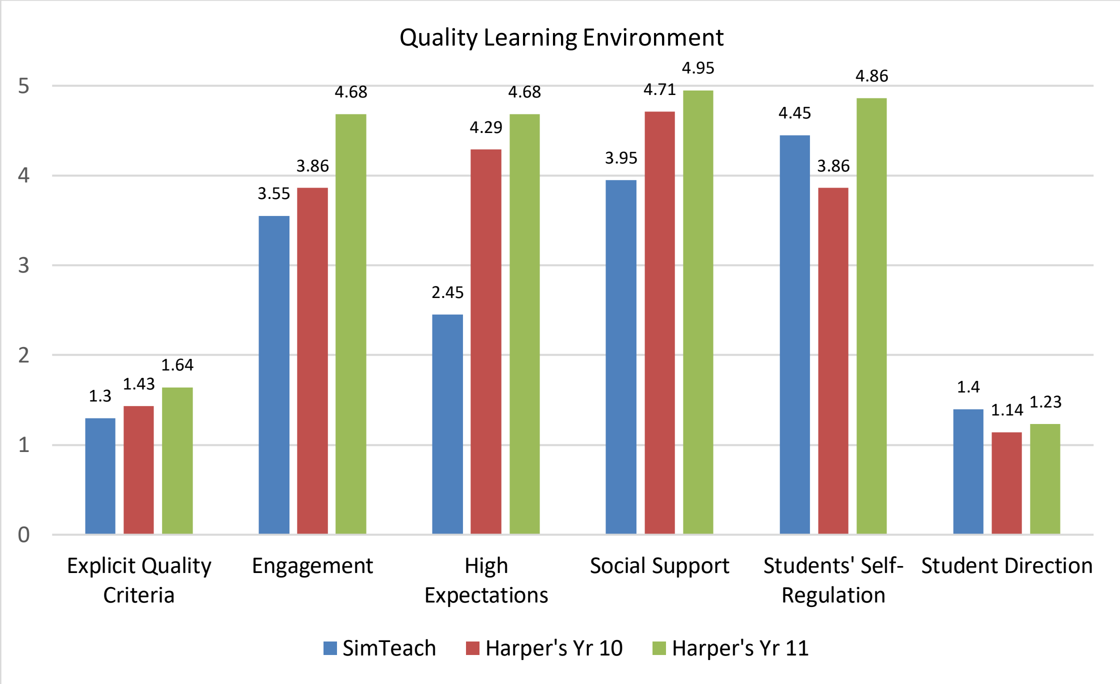

Figure 2. Comparison of the average codes awarded to the SimTeach lessons and Harper's (2023) Year 10 and 11 lessons for each element of the quality learning environment dimension of the QTM

Our results indicate an average code of 2.85 out of 5 for the quality learning environment, with particularly low codes in explicit quality criteria (1.3) and student direction (1.4). In contrast, we observed a strong performance in engagement (3.55), social support (3.95) and students' self-regulation (4.45). This suggests that while participants were effectively engaging with the content and supporting one another, they struggled with establishing clear quality criteria and allowing students to take an active role in their learning. Harper's (2023) results complement these findings, as they also reported lower results for explicit quality criteria (1.43 and 1.64) and student direction (1.14 and 1.23) across all classes (Figure 2).

However, Harper (2023) noted that engagement (4.68), high expectations (4.68), social support (4.95) and self-regulation (4.86) were consistently high, with Year 11 mathematics achieving the highest codes in these areas. On the other hand, in Year 10, engagement (3.86), high expectations (4.29), social support (4.71) and self-regulation (3.86) were also coded highly compared to explicit quality criteria (Figure 2). While both our results and Harper's (2023) findings indicate strengths in engagement and social support, they also highlight a critical need to enhance explicit quality criteria and student direction. This comparison advocates for strategies that can incorporate more student work, particularly in traditionally teacher-directed subjects like mathematics.

These findings highlight potential areas for pedagogical intervention within simulation-based teacher preparation. Incorporating prompts or scaffolding within SimTeach to encourage participants to clarify quality criteria and provide opportunities for student-led learning could enhance the transfer of these skills to real classroom settings. Understanding how cognitive load (Kirschner et al., 2018) and the constraints of a microteaching simulation influence participant performance can help refine both the design of simulated experiences and the interpretation of QTM scores. Overall, the discrepancies observed suggest that simulations provide valuable opportunities for teacher development but must be carefully structured to replicate the full complexity of student-centred learning environments.

Significance

In our results, significance emerged as the most difficult dimension for participants to incorporate into their lessons, with an average code of 1.93 out of 5, indicating that making learning meaningful and connecting it to broader contexts posed a challenge. In contrast, inclusivity stood out as a strength, with participants consistently receiving high codes (4.3), showing that diverse student engagement was well addressed. The results also showed moderately higher codes in background knowledge (2.45), suggesting participants were somewhat more effective in activating prior knowledge (Figure 3).

Harper's (2023) findings align with ours in terms of inclusivity, which was also high (5) across all classes they studied, indicating consistent student engagement. However, Harper's (2023) analysis dives deeper into subject-specific patterns. Mathematics focused more narrowly on internal connections within the discipline (strong classification; Bernstein, 2003), contributing to lower knowledge Integration codes (2.43) in mathematics except for Year 11, which was 3.05 (Figure 3). One critical distinction between the two studies is the treatment of cultural knowledge. Harper (2023) highlighted low codes across all classes, reflecting a gap in incorporating non-dominant or diverse knowledge systems in mathematics. Despite this emphasis on inclusivity (Harper, 2023), the absence of cultural knowledge suggests that implicit boundaries within the curriculum still limit diverse perspectives, particularly in westernised educational settings. Our data, by contrast, does not directly address cultural knowledge but similarly reflects the challenge of embedding broader, meaningful contexts into lesson planning, especially under the significance dimension. In both studies, cultural knowledge was coded as a 1 or slightly higher. Both our results and Harper's (2023) findings highlight the strengths of inclusivity and the challenges of integrating significance, but Harper's (2023) deeper analysis of subject-specific differences, particularly in how content is bound and the absence of cultural knowledge, provides an understanding of how curriculum design influences pedagogy and student engagement across disciplines.

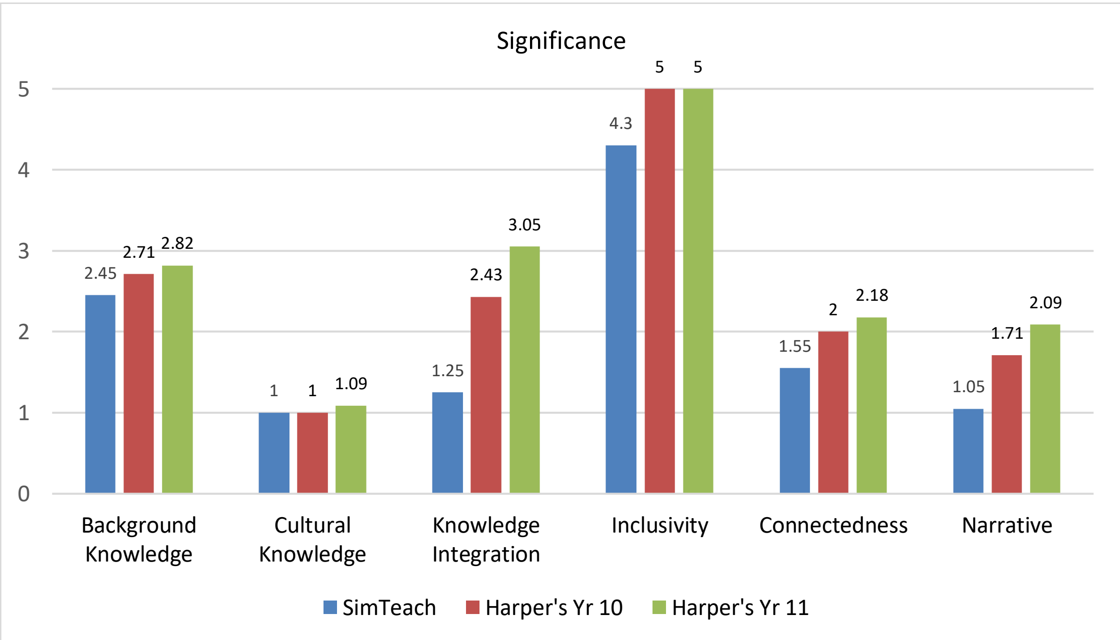

Figure 3. Comparison of the average codes awarded to the SimTeach lessons and Harper's (2023) Year 10 and 11 lessons for each element of the significance dimension of the QTM

These findings suggest that while participants can foster inclusivity and engage learners effectively, structural and curriculum-related factors, such as tightly bound content and limited cultural perspectives, constrain opportunities to enhance the significance of learning. Addressing these constraints with simulation could help participants develop strategies to connect disciplinary content to real-world and culturally diverse contexts.

In the next phase of our analysis, we compared the SimTeach observational results with those from five face-to-face class observations conducted in 2023 involving participants who were nearing the end of their postgraduate mathematics teacher education degree (Figure 4). While we anticipated an improvement in teaching quality at this stage of the programme, the results indicated that the codes remained similar during the face-to-face observations. This suggests that there may be specific areas where all mathematics teachers, regardless of their experience level, tend to struggle or excel due to the inherent nature of the subject, as was also observed in Harper's (2023) study.

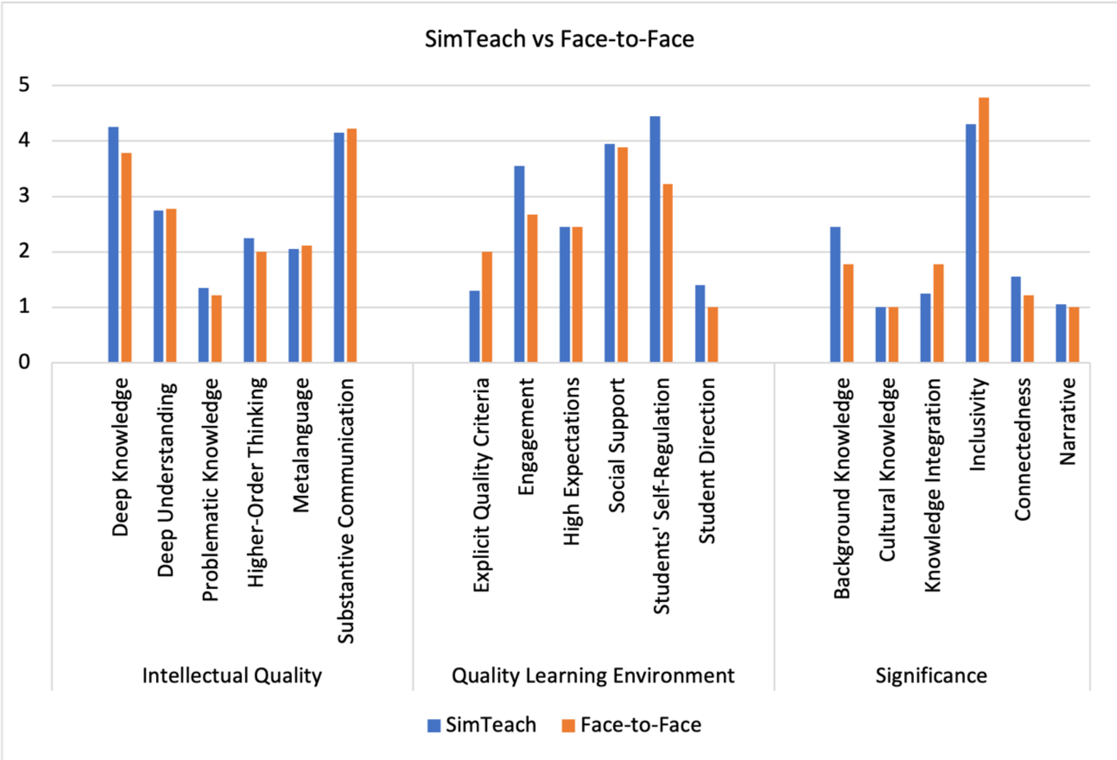

Figure 4. Comparison of the average codes awarded to the SimTeach lessons and face-to-face lessons for each element of the QTM

n comparing the coding between face-to-face and SimTeach lesson observations for the same cohort, we found that apart from students' self-regulation, most codes were quite similar. As the assumption of normality was violated for most elements as revealed by the *Shapiro-Wilk test* (Table 1), a *Mann-Whitney U* (Table 2) test was performed for each element of the QTM to determine whether the differences between face-to-face codes and SimTeach codes were significant.

Table 1

Descriptive data for each element of the QTM for both SimTeach and face-to-face observations and results of the Shapiro-Wilk test

QTM	Mode	Valid	Missing	Mean	Standard deviation	*Shapiro-Wilk*	P-value of *Shapiro-Wilk*	Minimum	Maximum
Deep knowledge	Face-to-face	5	0	4.000	1.414	0.767	0.042	2.000	5.000
	SimTeach	20	0	4.250	1.209	0.630	< .001	2.000	5.000
Deep understanding	Face-to-face	5	0	2.600	0.548	0.684	0.006	2.000	3.000
	SimTeach	20	0	2.750	1.118	0.849	0.005	1.000	4.000
Problematic knowledge	Face-to-face	5	0	1.000	0.000	NaN	NaN	1.000	1.000
	SimTeach	20	0	1.350	0.587	0.632	< .001	1.000	3.000
Higher-order thinking	Face-to-face	5	0	1.800	0.447	0.552	< .001	1.000	2.000
	SimTeach	20	0	2.250	1.209	0.814	0.001	1.000	4.000
Metalanguage	Face-to-face	5	0	2.200	0.837	0.881	0.314	1.000	3.000
	SimTeach	20	0	2.050	1.050	0.839	0.003	1.000	4.000
Substantive communication	Face-to-face	5	0	3.800	1.304	0.902	0.421	2.000	5.000
	SimTeach	20	0	4.150	1.040	0.745	< .001	2.000	5.000
Explicit quality criteria	Face-to-face	5	0	2.200	1.643	0.779	0.054	1.000	5.000
	SimTeach	20	0	1.300	0.733	0.477	< .001	1.000	4.000
Engagement	Face-to-face	5	0	2.800	1.095	0.828	0.135	1.000	4.000
	SimTeach	20	0	3.550	0.999	0.885	0.022	1.000	5.000
High expectations	Face-to-face	5	0	2.000	0.707	0.883	0.325	1.000	3.000
	SimTeach	20	0	2.450	1.432	0.835	0.003	1.000	5.000
Social support	Face-to-face	5	0	4.000	1.225	0.833	0.146	2.000	5.000
	SimTeach	20	0	3.950	0.945	0.851	0.006	2.000	5.000
Students' self-regulation	Face-to-face	5	0	3.200	1.483	0.956	0.777	1.000	5.000
	SimTeach	20	0	4.450	0.999	0.612	< .001	1.000	5.000
Student direction	Face-to-face	5	0	1.000	0.000	NaN	NaN	1.000	1.000
	SimTeach	20	0	1.400	0.940	0.483	< .001	1.000	5.000
Background knowledge	Face-to-face	5	0	1.800	0.837	0.881	0.314	1.000	3.000
	SimTeach	20	0	2.450	0.605	0.701	< .001	2.000	4.000
Cultural knowledge	Face-to-face	5	0	1.000	0.000	NaN	NaN	1.000	1.000
	SimTeach	20	0	1.000	0.000	NaN	NaN	1.000	1.000
Knowledge integration	Face-to-face	5	0	1.400	0.548	0.684	0.006	1.000	2.000
	SimTeach	20	0	1.250	0.550	0.522	< .001	1.000	3.000
Inclusivity	Face-to-face	5	0	4.600	0.548	0.684	0.006	4.000	5.000
	SimTeach	20	0	4.300	1.342	0.581	< .001	1.000	5.000
Connectedness	Face-to-face	5	0	1.200	0.447	0.552	< .001	1.000	2.000

QTM	Mode	Valid	Missing	Mean	Standard deviation	*Shapiro-Wilk*	P-value of *Shapiro-Wilk*	Minimum	Maximum
	SimTeach	20	0	1.550	0.605	0.737	< .001	1.000	3.000
Narrative	Face-to-face	5	0	1.000	0.000	NaN	NaN	1.000	1.000
	SimTeach	20	0	1.050	0.224	0.236	< .001	1.000	2.000

Table 2

p-values for the correlations between SimTeach and face-to-face observation codes in each element of the QTM

	W	p	Rank-biserial correlation	95% confidence interval [CI] for rank-biserial correlation	
				Lower	Upper
Deep knowledge	46.500	0.811	-0.070	-0.570	0.468
Deep understanding	45.000	0.750	-0.100	-0.590	0.444
Problematic knowledge	NaN[a]				
Higher-order thinking	43.500	0.667	-0.130	-0.609	0.419
Metalanguage	56.000	0.695	0.120	-0.427	0.603
Substantive communication	41.500	0.552	-0.170	-0.634	0.385
Explicit quality criteria	71.000	0.077	0.420	-0.129	0.772
Engagement	30.000	0.160	-0.400	-0.762	0.152
High expectations	45.500	0.776	-0.090	-0.583	0.452
Social support	53.500	0.831	0.070	-0.468	0.570
Students' self-regulation	21.500	0.035*	-0.570	-0.841	-0.070
Student direction	NaN[b]				
Background knowledge	27.500	0.093	-0.450	-0.786	0.092
Cultural knowledge	NaN[c]				
Knowledge integration	59.000	0.438	0.180	-0.376	0.640
Inclusivity	48.000	0.902	-0.040	-0.549	0.491
Connectedness	34.500	0.243	-0.310	-0.715	0.251
Narrative	NaN[d]				

Note. *Mann-Whitney U test*. Effect size is given by the rank-biserial correlation. W is the sum of ranks for one of the samples. p is the p-value, a probability indicating the strength of evidence against the null hypothesis of no difference between the groups. [a]The variance in problematic knowledge is equal to 0 after grouping on group; [b]student direction; [c]cultural knowledge; [d]narrative.

As shown in Table 2, it was found that there was no significant difference between groups in any element except for students' self-regulation ($p = 0.035$), which was higher in the SimTeach group than the face-to-face group ($r_{rb} = -0.570$, 95% CI [-0.841, -0.070]). This indicates that the shift from the online SimTeach environment to face-to-face classrooms did not significantly affect the teaching quality across the elements of the QTM. While self-regulation was coded slightly higher in the SimTeach lessons, it's important to consider the fact that the SimTeach students are not real students, though they behaved similarly. The difference, however, is not substantial enough to invalidate the use of SimTeach lessons for assessing students' self-regulation. The fact that both environments produced comparable results supports the validity of using SimTeach lessons as an alternative to face-to-face observations for evaluating teaching quality. However, we would argue that while teaching quality is measurable through SimTeach lessons, understanding students' progress may still require face-to-face observations, especially when educators and researchers need to observe specific student groups.

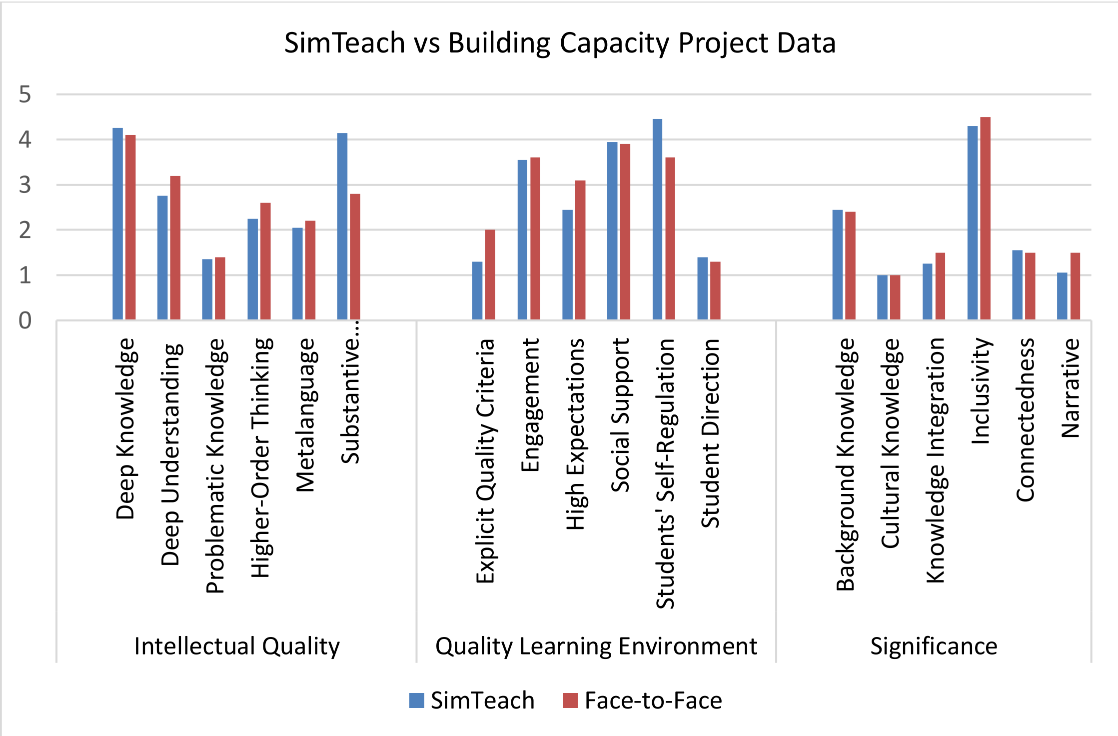

Figure 5. Comparison of the codes awarded to the SimTeach lessons and face-to-face lessons conducted as part of the Building Capacity for Quality Teaching in Australian Schools initiative for each element of the QTM

In the final phase of our analysis, we compared the results of our SimTeach observations with those of the observations conducted as part of the Building Capacity for Quality Teaching in Australian Schools initiative (Gore et al., 2023). As can be seen in Figure 5, the average codes between the SimTeach observations and the face-to-face observations were similar. The most significant difference that was not observed in the previous comparisons is in substantive communication, with the SimTeach observations receiving an average code of 4.15 and the face-to-face observations in this project receiving a code of 2.8. Other elements with notable differences included explicit quality criteria (1.3 for SimTeach and 2 for face-to-face) and students' self-regulation (4.45 for SimTeach and 3.6 for face-to-face). The differences between these two elements align with the differences observed between the SimTeach and the face-to-face observations.

These discrepancies highlight how environmental and contextual factors including social interaction, classroom complexity and immediacy of feedback can influence QTM scores. Understanding these differences is critical for interpreting simulation-based observations and for refining SimTeach as a tool for teacher preparation. As a simulation-based learning tool, it provides pre-service teachers with immersive, interactive experiences that extend beyond mere observation, supporting the development of reflective practice, classroom management strategies and decision-making skills in a controlled environment. Furthermore, SimTeach has the potential to align with emerging trends in educational technology, such as AI-driven feedback, virtual simulations and scalable teacher training platforms. If integrated thoughtfully into higher education, tools like these could enhance access to authentic teaching experiences, support more equitable learning opportunities and contribute to the ongoing digital transformation of teacher education.

Conclusion

Despite initial challenges during coding, particularly within the SimTeach observation context where student responses and facial expressions were crucial, we found that simulated lessons produced results closely aligned with those from face-to-face data. By comparing SimTeach coding with traditional mathematics lesson coding, we observed minimal differences in key areas, such as deep understanding and explicit quality criteria, suggesting that the mode of lesson delivery may not significantly affect these aspects of teaching quality. However, common struggles remain, particularly in student direction and cultural knowledge, where mathematics teachers may feel pressured to maintain lesson control, limiting student autonomy and cultural context integration. This challenge aligns with Wager (2012)'s findings that mathematics teachers often struggle to incorporate students' real-life mathematical knowledge and cultural practices into lessons due to the difficulty of identifying relevant scenarios and potential misunderstandings around cultural practices.

Our results show that while SimTeach participants demonstrated strengths in most elements of the QTM. When compared to Harper's (2023) data, SimTeach participants showed less strength in metalanguage, high expectations and knowledge integration but outperformed Harper's observations in students' self-regulation. They also showed better substantive communication and self-regulation than the building capacity project data.

It is essential to acknowledge our study's limitations, particularly the small sample size, the limited access to raw data and the difficulty of coding some elements in a simulated environment, such as deep understanding and explicit quality criteria. We were unable to access the raw data from Harper's (2023) study or the Building Capacity for Quality Teaching in Australian Schools project (Gore et al., 2023), resulting in our findings lacking statistical depth. Therefore, we emphasise the need for further research and targeted professional development in mathematics education to support teachers in integrating problematic knowledge, explicit quality criteria and elements of significance, especially cultural knowledge and student direction, across both simulated and real-world classrooms. Addressing these areas can lead to more engaging and culturally responsive teaching practices, ultimately enhancing the quality of mathematics education in both virtual and face-to-face settings.

In conclusion, we argue that virtual teaching simulations offer a promising alternative, or a valuable complement, to traditional classroom observations in assessing teacher quality. The findings have broader implications for supporting the digital transformation of teacher education. First, simulation offers scalable, flexible and immersive learning experiences that complement traditional in-person methods. Second, it enables the evaluation of teaching practices in controlled environments, providing consistent and comparable feedback across contexts. Third, attention to technological infrastructure, accessibility and inclusive design is essential to ensure equity for all preservice teachers. Theoretically, these results affirm the value of simulation-based learning and immersive technologies in teacher education, showing how virtual experiences can reinforce core teaching competencies through innovative assessment approaches. The authors advocate for policies and guidelines that promote the ethical, equitable and effective implementation of simulation to fully realise its pedagogical potential.

Author contributions

Sabrina Syed: Conceptualisation, Investigation, Writing – original draft, Writing – review and editing; **Claire Bates:** Data curation, Investigation, Formal analysis, Writing – review and editing; **Elena Prieto-Rodriguez:** Writing – review and editing; **Susan Ledger:** Review and editing.

Acknowledgements

We sincerely appreciate the cooperation of the research participants who took part in the Mathematics Retraining Programme, funded by the NSW Department of Education. We would also like to extend our gratitude to our colleague, Ms. Beth Preston, for her valuable contributions as a key member of our SimTeach lessons.

References

Ade-Ojo, G. O., Markowski, M., Essex, R., Stiell, M., & Jameson, J. (2022). A systematic scoping review and textual narrative synthesis of physical and mixed-reality simulation in pre-service teacher training. *Journal of Computer Assisted Learning*, *38*(3), 861–874. https://doi.org/10.1111/jcal.12653

Allen, D. W., & Clark, R. J. (1967). Microteaching: Its rationale. *The High School Journal*, *51*(2), 75–79. http://www.jstor.org/stable/40366699

Amedu, J., & Hollebrands, K. F. (2022). Teachers' perceptions of using technology to teach mathematics during COVID-19 remote learning. *REDIMAT*, *11*(1), 71–85. https://doi.org/10.17583/redimat.8872

Berg, C., Dieker, L., & Scolavino, R. (2023). Using a virtual avatar teaching simulation and an evidence-based teacher observation tool: A synergistic combination for teacher preparation. *Education Sciences*, *13*(7), 744. https://www.mdpi.com/2227-7102/13/7/744

Bernstein, B. (2003). On the classification and framing of educational knowledge. In *Theoretical Studies Towards a Sociology of Language* (Vol. I, pp. 156–177). Routledge. https://doi.org/10.4324/9780203014035-14

Blazar, D., Gilbert, B., Herlihy, C., & Gogolen, C. (2018). Using tablets to explore the potential for video-based classroom observations for research and professional development. *International Journal of Education In Mathematics, Science and Technology*, *6*(2), 122–135. https://www.learntechlib.org/p/190452/

Brucklacher, B., & Gimbert, B. (1999). Role-playing software and WebQuests-what's possible with cooperative learning and computers. *Computers in the Schools*, *15*(2), 37–48. https://doi.org/10.1300/J025v15n02_05

Cohen, J., & Goldhaber, D. (2016). Building a more complete understanding of teacher evaluation using classroom observations. *Educational Researcher*, *45*(6), 378–387. https://doi.org/10.3102/0013189x16659442

Cooper, J., & Allen, D. (1970). Microteaching: History and present status. ERIC. https://files.eric.ed.gov/fulltext/ED036471.pdf

Council of Chief State School Officers. (n.d.). *InTASC model core teaching standards and learning progressions for teachers*. https://learning.ccsso.org/intasc-model-core-teaching-standards-and-learning-progressions-for-teachers

Dittrich, L., Aagaard, T., & Hjukse, H. (2022). The perceived affordances of simulation-based learning: Online student teachers' perspectives. *International Journal of Educational Technology in Higher Education*, *19*(1), 60. https://doi.org/10.1186/s41239-022-00366-2

Ferri, F., Grifoni, P., & Guzzo, T. (2020). Online learning and emergency remote teaching: Opportunities and challenges in emergency situations. *Societies*, *10*(4), 86. https://doi.org/10.3390/soc10040086

Fischetti, J., Ledger, S., Lynch, D., & Donnelly, D. (2022). Practice before practicum: Simulation in initial teacher education. *The Teacher Educator*, *57*(2), 155–174. https://doi.org/10.1080/08878730.2021.1973167

Gillett-Swan, J. (2017). The challenges of online learning: Supporting and engaging the isolated learner. *Journal of Learning Design*, *10*(1), 20–30. https://doi.org/10.5204/jld.v9i3.293

Gore, J., Lloyd, A., Smith, M., Bowe, J., Ellis, H., & Lubans, D. (2017). Effects of professional development on the quality of teaching: Results from a randomised controlled trial of Quality Teaching Rounds. *Teaching and Teacher Education*, *68*, 99–113. https://doi.org/10.1016/j.tate.2017.08.007

Gore, J., Miller, D., Fray, L., & Patfield, S. (2023). *Building capacity for quality teaching in Australian schools 2018–2023* (Final report). University of Newcastle.

https://www.newcastle.edu.au/__data/assets/pdf_file/0008/956231/Building-Capacity-for-Quality-Teaching-in-Australian-Schools-Final-Report-2023.pdf

Gore, J. M., Miller, A., Fray, L., Harris, J., & Prieto, E. (2021). Improving student achievement through professional development: Results from a randomised controlled trial of quality teaching rounds. *Teaching and Teacher Education*, *101*, 103297. https://doi.org/10.1016/j.tate.2021.103297

Harper, M. D. (2023). Demystifying subjects, troubling status: A pedagogical analysis of high school mathematics and drama in the Australian schooling context [Doctoral dissertation, University of Newcastle]. University of Newcastle Research Repository. https://hdl.handle.net/1959.13/1503980

Heafner, T. L., Petty, T. M., & Hartshorne, R. (2011). Evaluating modes of teacher preparation: A comparison of face-to-face and remote observations of graduate interns. . *Journal of Digital Learning in Teacher Education*, *27*(4), 154–164. https://doi.org/10.1080/21532974.2011.10784672

Howard, A. J. (2010). Teacher appraisal: The impact of observation on teachers' classroom behaviour [Doctoral dissertation, University of Warwick]. University of Warwick Institutional Repository. http://go.warwick.ac.uk/wrap/3728

Hsiao, J.-C., Chen, S.-K., Chen, W., & Lin, S. S. J. (2022). Developing a plugged-in class observation protocol in high-school blended STEM classes: Student engagement, teacher behaviors and student-teacher interaction patterns. *Computers & Education*, *178*, 104403. https://doi.org//10.1016/j.compedu.2021.104403

Kirschner, P. A., Sweller, J., Kirschner, F., & Zambrano R, J. (2018). From cognitive load theory to collaborative cognitive load theory. *International Journal of Computer-Supported Collaborative Learning*, *13*, 213–233. https://doi.org/10.1007/s11412-018-9277-y

Ladwig, J. G. (2009). Working backwards towards curriculum: On the curricular implications of Quality Teaching. *The Curriculum Journal*, *20*(3), 271–286. https://doi.org/10.1080/09585170903195886

Ledger, S. (2019). Simulation in higher education: Choice, challenges and changing practice. In *Proceedings of the 3rd International Conference on Education Innovation* (pp. 387–392). https://doi.org/10.2991/icei-19.2019.91

Lew, S., Gul, T., & Pecore, J. L. (2021). ESOL pre-service teachers' culturally and linguistically responsive teaching in mixed-reality simulations. *Information and Learning Sciences*, *122*(1-2), 45–67. https://doi.org/10.1108/ILS-01-2020-0012

Liang, J. (2015). Live video classroom observation: An effective approach to reducing reactivity in collecting observational information for teacher professional development. *Journal of Education for Teaching*, *41*(3), 235–253. https://doi.org/10.1080/02607476.2015.1045314

Lindberg, S., & Jönsson, A. (2023). Preservice teachers training with avatars: A systematic literature review of "human-in-the-loop" simulations in teacher education and special education. *Education Sciences*, *13*(8), 817. https://www.mdpi.com/2227-7102/13/8/817

Merrett, F. (2006). Reflections on the Hawthorne effect. *Educational Psychology*, *26*(1), 143–146. https://doi.org/10.1080/01443410500341080

Miltenberger, R. G. (2016). *Behavior modification: Principles and* procedures (6th ed.). Cengage Learning.

Mukuka, A., & Alex, J. (2024). Review of research on microteaching in mathematics teacher education: Promises and challenges. *Eurasia Journal of Mathematics, Science and Technology Education*, *20*(1), em2381. https://doi.org/10.29333/ejmste/13941

Newmann, F. M. (1996). Authentic achievement: *Restructuring schools for intellectual quality*. Jossey-Bass Publishers.

New South Wales Department of Education and Training. (2003). *Quality teaching in NSW public schools: A classroom practice guide.* Professional Learning and Leadership Development Directorate. https://education.nsw.gov.au/content/dam/main-education/en/home/teaching-and-learning/professional-learning/quality-teaching-rounds/QT_Classroom_Practice_Guide.pdf

Patfield, S., Gore, J., Prieto, E., Fray, L., & Sincock, K. (2025). Towards quality teaching in higher education: Pedagogy-focused academic development for enhancing practice. *International Journal for Academic Development*, *30*(4), 553–568. https://doi.org/10.1080/1360144X.2022.2103561

Quality Teaching Academy. (2020). *Quality teaching classroom practice guide* (3rd ed.). State of NSW, Department of Education.

Richards, K. (2003). *Qualitative inquiry in TESOL*. Palgrave Macmillan UK. https://doi.org/10.1057/9780230505056

Rosati-Peterson, G. L., Piro, J. S., Straub, C., & O'Callaghan, C. (2021). A Nonverbal immediacy treatment with pre-service teachers using mixed reality simulations. *Cogent Education*, *8*(1), 1882114. https://doi.org/10.1080/2331186X.2021.1882114

Taylor, M. (2024). Accessibility and inclusivity in online education: A comparative study of policies and practices in Australia. *International Journal of Online and Distance Learning*, *5*(1), 25–36. https://doi.org/10.47604/ijodl.2440

Turoff, M. (1995). Designing a virtual classroom. *International Journal of Educational Telecommunications*, 1(2), 245–262. https://www.learntechlib.org/p/15162

University of Newcastle. (2023). SimTeach. Retrieved 24 April 2023, from https://www.newcastle.edu.au/school/education

Wager, A. A. (2012). Incorporating out-of-school mathematics: From cultural context to embedded practice. *Journal of Mathematics Teacher Education*, *15*, 9–23. https://doi.org/10.1007/s10857-011-9199-3

Waxman, H. C., Tharp, R. G., & Hilberg, R. S. (Eds.). (2004). Observational research in US classrooms: New approaches for understanding cultural and linguistic diversity. Cambridge University Press.

Whitcomb, S. A., & Merrell, K. W. (2013). *Behavioral, social, and emotional assessment of children and adolescents* (4th ed.). Routledge. https://doi.org/10.4324/9780203818596

Yohannes, Y., Juandi, D., Diana, N., & Sukma, Y. (2021). Mathematics teachers' difficulties in implementing online learning during the COVID-19 Pandemic. *Journal of Hunan University Natural Sciences*, *48*(5), 87–98. https://www.jonuns.com/index.php/journal/article/view/581

Ziebell, N., Acquaro, D., Seah, W. T., & Pearn, C. (2020). *Australian education survey: Examining the impact of COVID-19* [Report Summary]. The University of Melbourne.

Corresponding author: Sabrina Syed, ssyed@csu.edu.au

Please cite as: Syed, S., Bates, C., Prieto-Rodriguez, E., & Ledger, S. (2025). Evaluating the validity of SimTeach lessons as an alternative to face-to-face classroom observations. *Australasian Journal of Educational Technology*, *41*(6), 18–35. https://doi.org/10.14742/ajet.10536