



The Interactive Virtual Training for Teachers (IVT-T) to Practice Classroom Behavior Management

A. Delamarre^{*,a}, E. Shernoff^b, C. Buche^c, S. Frazier^d, J. Gabbard^e, C. Lisetti^a

^a VISAGE Lab, Knight Foundation School of Computing and Information Sciences, Florida International University, USA

^b Department of Psychology, Rutgers University, USA

^c CROSSING, ENIB, CNRS IRL 2010, Australia

^d Department of Psychology, Florida International University, USA

^e Grado Department of Industrial and Systems Engineering, Virginia Tech, USA

ARTICLE INFO

Keywords:

Human-centered Interaction design
Virtual reality (VR)
User experience (UX)
Usability
Requirements
Teachers Training

ABSTRACT

In this article, we discuss the 3D Interactive Virtual Training for Teachers system (IVT-T) which we built for teachers to hone in their classroom management skills. By interacting with thirty fully autonomous and animated realistic 3D virtual students in two detailed and realistic 3D classrooms, teachers can practice their behavior management and receive feedback and reflection opportunities on their performances. Lack of experience in managing students' disruptive behavior in classrooms is one of the main causes of teacher turnover and IVT-T provides them with such an experience in virtual reality using 3D graphics and spatialized sounds. IVT-T is a low-cost stand-alone fully autonomous system that teachers can interact with on their laptop, without special equipment or personnel to operate the system. Although previous training systems for teachers have been successfully developed, requirements to guide future research and development of such systems have not been established. We present a set of six requirements (Behavioral Fidelity, Environment Fidelity, Instructional Design, Autonomy, Interactivity, and Scalability) for VTEs for teachers established from our survey of the literature and from our iterative lifecycle activities to build IVT-T. We describe IVT-T's modular architecture designed to meet these VTE requirements and to enable the addition of an integrated intelligent tutor. We evaluated IVT-T to assess whether all requirements have been met. The evaluation of the behavioral and graphical fidelity of our virtual students conducted by education experts using mixed methods indicated that IVT-Ts graphics and students' behaviors were very *realistic*. The evaluation of IVT-T's User Interface (UI) usability confirmed IVT-T's *effectiveness*, *efficiency*, and *learnability*. We present current findings supporting the *soundness* of IVT-T instructional design. We conclude with implications of our proposed requirements for future VTEs for learning social skills, and our future plans for IVT-T.

1. Introduction

Disruptive behavior in the classroom is the main cause of stress for teachers, particularly for teachers with less than five years of experience [Shernoff et al. \(2016, 2011\)](#). Limited training on how to deal with disruptive students is one of the greatest contributors to new teacher turnover [Ingersoll and Smith \(2003\)](#); [Shernoff et al. \(2016\)](#). Given the high demand for qualified teachers, there is an urgent need to support teacher training by tackling teachers' main issues: (1) teachers are unprepared for the realities of teaching [Grossman and McDonald \(2008a\)](#); (2) teachers have few opportunities to practice while receiving expertly

tailored feedback about their mistakes [Denton and Hasbrouck \(2009\)](#); [Shernoff et al. \(2015\)](#); and (3) teachers have few opportunities for reflecting on their skills (or lack of) and how to resolve problems [Merrill \(2009\)](#).

Teachers' ability to prevent and manage behavior problems directly contributes to student success and students learning outcomes, especially for students with learning difficulties and students at risk for emotional and behavioral disabilities [Oliver and Reschly \(2010\)](#). Prevention and management of disruptive behaviors promotes student success by increasing instruction effectiveness, and maximizing learning opportunities [Creemers \(1994\)](#); [Crone and Teddlie \(1995\)](#); [Oliver and](#)

* Corresponding author. Tel.: +7868570062.

E-mail addresses: adela177@fiu.edu (A. Delamarre), elisa.shernoff@rutgers.edu (E. Shernoff), cedric.buche@enib.fr (C. Buche), slfrazi@fiu.edu (S. Frazier), jgabbard@vt.edu (J. Gabbard), lisetti@cs.fiu.edu (C. Lisetti).

<https://doi.org/10.1016/j.ijhcs.2021.102646>

Received 9 June 2020; Received in revised form 21 March 2021; Accepted 30 March 2021

Available online 16 April 2021

1071-5819/© 2021 Elsevier Ltd. All rights reserved.

Reschly (2007). Most of the time, teachers go through a trial-and-error approach in real classrooms to improve their skills in behavior management, resulting in teachers and students' having a negative learning experience, uncomfortable classroom climate, and strained interpersonal relationships. Supporting the learning of classroom behavior management (CBM) skills can mitigate negative classroom interactions, and can lead to facilitation of student learning.

Over the past two decades, simulation-based VTEs (training systems which include real world simulations in a virtual environment) have been applied in a variety of domains, e.g. fire-fighter training, army soldier training, procedural training, safety training, and risk environment training Barot et al. (2013); Gerbaud et al. (2008); Nakayama et al. (2015); Nakhhal et al. (2016); Querrec et al. (2004). VTEs offer many advantages compared to traditional training: VTEs can simulate conditions for situations that are impossible, dangerous, or too costly to reproduce (e.g. piloting a plane pilots, responding to dangerous chemical accidents); VTEs act as a sandbox where errors committed inside the virtual environments have no impact on reality, and allow users to re-iterate the training until goals are achieved; and VTEs support active learning exposing users to situations requiring their intervention, which provides a hands-on experience.

VTEs are therefore a viable solution for the issues faced by teachers, and VTEs designed to help teachers develop CBM skills have been used successfully Barmaki and Hughes (2015); Dieker et al. (2015); Gregory et al. (2013); Gupta et al. (2008); Hayes et al. (2013b); Kelleci and Aksoy (2020); Kervin et al. (2006); Nakhhal et al. (2016); Straub et al. (2014); Tichon (2007); Zibit and Gibson (2005). VTEs offer a psychologically safe environment where mistakes managing virtual students in the classroom have no impact on real students. VTEs can also help systematically monitor teachers' training, and VTEs can provide feedback on teachers' performance with the virtual students.

Building VTEs for learning and enhancing social skills, however, still presents many scientific and technical challenges, including: 1) identifying a set of functional and user experience (UX) requirements needed for the design and implementation of effective and engaging training systems for teachers; 2) designing a system with a modular architecture so that it can be expanded upon; 3) deciding upon the appropriate level of graphical realism of the 3D virtual environment and students, while weighing factors such as costs and performance; 4) deciding whether to program the virtual students to be fully automated so that users can use the system anytime or whether to require an instructor to tele-operate the virtual students during interactions (behind users' view); 5) finding the right evaluation methodologies to assess the numerous and varied aspects of the system and its UIs; 6) integrating features to maintain users' engagement with the system by borrowing from game design elements; among other challenges.

In our current work, we build upon previous studies on VTEs for teachers, and on traditional professional development in CBM Shernoff et al. (2018) to identify key features required for the development of an effective classroom simulator to support teachers' training. Based on these features, we introduce the Interactive Virtual Training for Teachers (IVT-T), a 3D VTE which simulates 30 virtual students evolving in 2 virtual classrooms (15 1st grade students and 6th grade students), to provide teachers with CBM practice. IVT-T also offers constructive feedback and reflection opportunities about situations faced during the practice sessions, features which are not always included in the design of VTEs for teachers, yet essential to better understand CBM.

In this article, Section 2 provides a brief overview of the human-centered lifecycle that our team of computer scientists and education experts undertook over the past four years to create increasingly more sophisticated IVT-T prototypes, up to the IVT-T current version which we present. Section 3 presents a set of general requirements which our survey of the literature and our own lifecycle revealed need to be considered, and which will prove useful for researchers on VTEs for teachers.

Section 4 describes the specifics of IVT-T's architecture, the design of

IVT-T's thirty 3D virtual students and two classrooms, the IVT-T simulator responsible for playing the scenarios within the virtual classrooms, and IVT-T's UI for teachers (the second UI we built for IVT-T administrators to gather user data and conduct analyses is not described).

In Section 5, we present the various evaluations of IVT-T that we conducted using mixed methods to assess IVT-T graphical and behavioral components realism, the evaluation of IVT-T usability, as well as our latest findings on the use of IVT-T by early career teachers (ECTs). It is important to note that although IVT-T has all the features of a training system for teachers, our focus on this was to consider ECTs as end-users, because research indicates that ECTs are most likely to leave teaching because of lacking CBM experience Atteberry et al. (2017); Ingersoll and Strong (2011). Lastly, in Section 6, we discuss the implications of the proposed requirements for future VTEs for social skills, and our plans to evaluate the impact IVT-T has on ECTs' transfer of learning in real classrooms.

2. Overview of IVT-T Context and Development Lifecycle

2.1. Classroom Behavior Management Strategies in IVT-T

Although IVT-T can be useful to any teachers interested in improving on CBM, the goal of IVT-T is to provide ECTs with realistic classroom teaching experience. IVT-T offers low-stakes training and maximize active learning opportunities including reflection, and feedback. The difficulty of managing classroom behavior is exacerbated for ECTs who already receive limited mentoring in behavior management Grossman and McDonald (2008b), and must acquire these skills on-the-job with real students while delivering instruction. Fast-paced, high-stakes, live instruction leaves little time for practice with feedback, which can be costly to teachers and learners Henry et al. (2011); Schussler et al. (2017).

IVT-T is designed to provide experiential training in parallel with a didactic 8-week course with specific learning outcomes. During the course, ECTs learn about the following concepts and during their practice with IVT-T, ECTs learn to identify them:

- antecedent-behavior-consequence (ABC) cycles Kazdin (2008),
- positive classroom climate,
- proactive monitoring, and
- effective redirection.

In an ABC cycle, antecedents describe what occurs before the behaviors and what will influence the behaviors (e.g. instructions, gestures, looks from peers). Behaviors are the actions that the individual actually does or does not do, and consequences characterize what follows the behaviors, which will eventually increase, decrease, or have no impact on the individual's behaviors. Any interchange is an ongoing sequence of antecedents-behavior-consequences, with sequences always starting with an antecedent Kazdin (2008). Classroom climate refers to attitudes, standards and tone used by teacher and students in a classroom. A positive classroom climate feels safe, respectful, welcoming, and supportive of student learning. Proactive monitoring consist of identifying early cues of disruptive behaviors and effective redirection involves efficient, early and private redirection, combined with a consequence hierarchy as well as praises for student's compliance.

As shown in Fig. 1, IVT-T scenarios were written so as to provide exposure to situations where these concepts can be experienced virtually. Each scenario starts with an opening scene describing the situation, for instance describing what happened the day before the scenario is taking place, to situate ECTs in the narrative of the scenario. Scenarios were constructed with ABC cycles, providing the teacher trainee with opportunities to identify them and to choose strategies of various levels of efficiency. The action consequence options presented to ECTs were implemented based on the evidence-based strategies, which mitigate disruptive behaviors and enhance student's attention, compliance, and

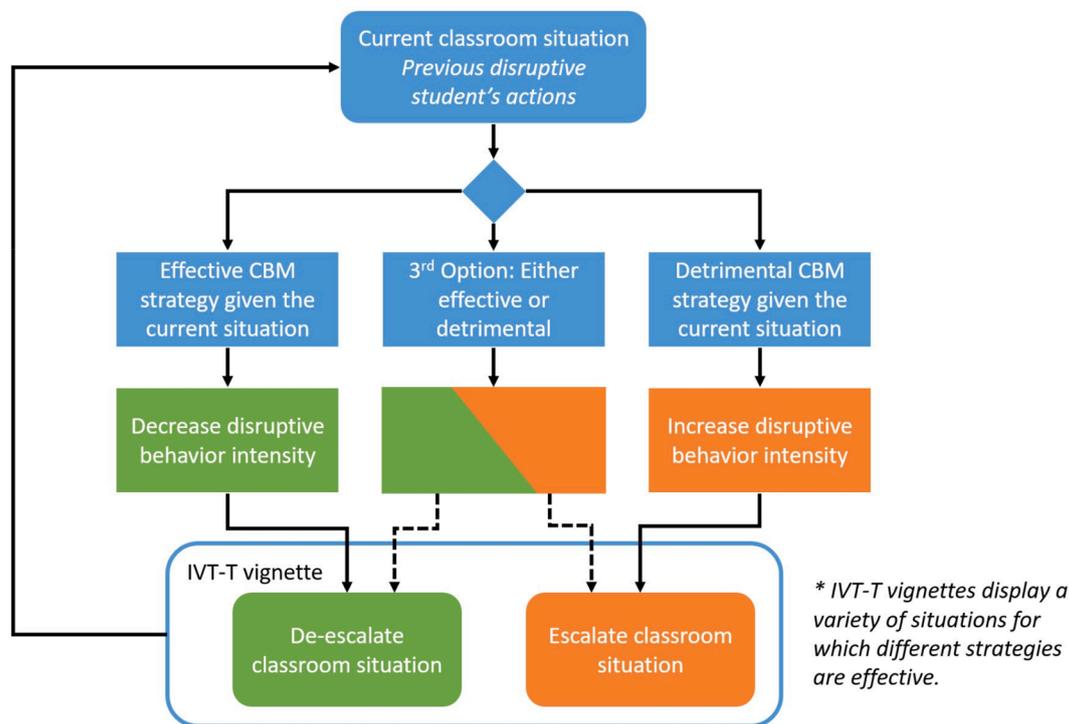


Fig. 1. Evolution of the disruptive behaviors based on the context of the vignette and on the strategies selected by trainees.

engagement Evertson and Weinstein (2013); Junod et al. (2006); Kazdin (2008), include the following: *praise, ignore, redirect, use of proximity, instructions, empathy, and if/then statements.*

Depending on the student behavior and on the situation, resorting to effective strategies deescalate the disruptive behaviors (student becomes more engaged and compliant), whereas detrimental strategies escalate the situation (student becomes more aggressive or more off-task). These interactions simulate real classroom antecedent-behavior-consequence cycle of disruptive behaviors Kazdin (2008).

For example, [the teacher says: "Jordan, that's inappropriate. You've lost a point."] composes the *antecedent* of the behavior [Jordan says: "It's not fair. I was answering your question!"]. At this point, multiple potential *consequences* to that behavior are provided as choices the teacher needs to select from, where a consequence will either escalate or deescalate the situation. An example of a deescalating consequence is [the teacher says in an encouraging tone: "Actually, we're going to talk about how ratios relate to batting averages. The sooner you complete the Do Now, the sooner we can talk about baseball."]. Examples of escalating consequence options are [the teacher says in an irritated tone: "It's not fair that you come in and disrupt my class every day, either."], or [the teacher says in a firm tone: "Your language was inappropriate. I need you to get started now."]. If the trainee chooses the escalating consequences, the student reacts with a *behavior* with increased disruption, e.g. [Jordan stops working and cooperating completely]. In either case, the student's behavior leads to the current classroom situation, which in turn becomes the antecedent for the next ABC cycle.

The scenarios also offer opportunities to practice *proactive monitoring* (e.g., identifying that the student is trying to get attention by humming), *effective redirection* (e.g., asking the student to lead the review for the class in order to stop the humming and involves the student with the group), and *identifying positive classroom climate* (e.g., the student is writing on the board for the class).

2.2. IVT-T Initial Requirements

The IVT-T's education experts on our team originally specified that: (1) IVT-T should present highly realistic graphics and behaviors of

classroom and virtual students, raising a requirement for graphical and behavioral realism; (2) ECTs should be able to practice autonomously and repeatedly at any time their behavior management skills in realistic classroom situations, pointing to a low personnel requirement so that no instructor is needed to run the training sessions; and (3) IVT-T should be accessible to ECTs at any time, emphasizing the importance of online access and low-technology requirements (laptop or desktop computer).

2.3. IVT-T Lifecycle Overview

Our main research objectives for IVT-T were to:

- identify what are the main *requirements* for building effective, usable, and enjoyable VTE for teachers (VTE-Ts);
- design and implement IVT-T so that the system is highly *usable* by novice and non-technical users and so that it provides *realistic* classroom situations that users find authentic;
- assess IVT-T *fidelity* in terms of whether it is used as intended; and
- assess IVT-T *feasibility* in terms of transfer of knowledge and skills from the virtual classroom to ECTs live classrooms.

In this article we discuss how we reached our first three main objectives which led us to build and validate IVT-T realism, usability, and usage. Our fourth objective will be conducted by the education experts on our team and is outside the scope of this article.

Given the high emphasis on the usability of the system by our end-users - ECTs without technical skills - we adopted Hartson's user-centered iterative interaction design lifecycle har (2019). IVT-T development lifecycle is depicted in Fig. 2, showing how IVT-T underwent yearly evaluations over four consecutive years, leading to prototypes IVT-T 1.0 to IVT-T 4.0 of increasingly higher fidelity.

During the first cycle, discussions with the team of experts in education combined with our review of the literature helped us extend and refine the initial set of basic requirements for the development of virtual environment to support teacher training, which we describe in details in Section 3.

Given the emphasis on providing highly realistic visuals, the 3D

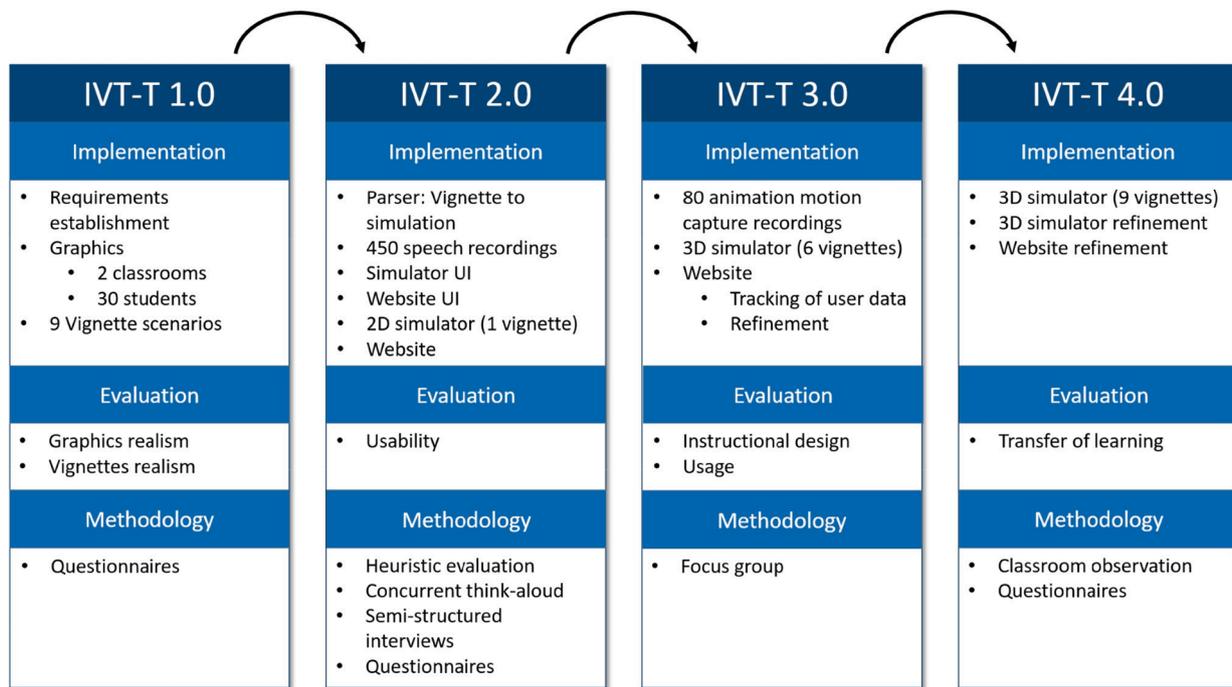


Fig. 2. IVT-T development lifecycle.

computer graphics were also developed during the first year for IVT 1.0: two virtual classrooms (6th grade and 1st grade) and thirty unique virtual students (15 1st graders and 15 6th graders) were created. As indicated in Fig. 2, the evaluation of the graphical classrooms and students was conducted using questionnaires. A board of experienced teachers provided feedback to improve the 3d models which were refined accordingly and iteratively (6 cycles of evaluation and refinements for the classroom and 4 cycles for the students) (discussed in details in Section 5.2). In parallel to the computer science graphics lifecycle, the education experts on the team had a similar lifecycle to develop and validate the scenario vignettes in terms of the realism of the disruptive students' behaviors Shernoff et al. (2018).

During the second cycle, we implemented the first versions of both the simulator, and the website UI. To enable the early evaluation of IVT-T UI and of the content of the scenarios (without having to wait for the time consuming generation of fine-tuned 3D graphics animations), we built IVT-T 2.0 as a hybrid prototype. In IVT 2.0 we implemented the main functionalities of the website, and the simulation of the scenarios were prototyped as a storyboard simulator: users interacted with a selection of vignette scenarios through a sequence of still images representing the final 2D version of the virtual students, placed in their desired position in the classroom. This allowed for the users to experience IVT-T UI as they were asked to complete main benchmark tasks we had identified, the only difference with a full blown prototype being that users had to click through the simulation storyboard pages, instead of seeing the simulation play automatically for them in the 3D classroom. As mentioned in Fig. 2, usability questionnaires, concurrent think-aloud (CTA) protocol and semi structured interviews were used to collect data from education majors, as described in details in Section 5.3.

During the third cycle, 80 animations were recorded using motion capture and integrated into IVT-T 3.0. A partial list of animations is provided in Table 2. Moreover, a parser was implemented to translate the vignette scenarios established by the education experts into 3D simulations. The parser allowed the efficient implementation of 5 more vignette scenarios. The IVT-T 3.0 website included features for the tracking of user data so that users can visualize their progress, and so that IVT-T education experts can track usage and measure the efficiency of the system. IVT-T 3.0 was evaluated in terms of its usage by ECTs from

high-poverty school in a North-Eastern state, as described in Section 5.4. The current IVT-T 4.3 version contains 9 vignette scenarios. Refinements of the simulator and of the website were made according the results obtained from the evaluation of previous IVT-T versions. The evaluation of the system in terms of transfer of knowledge and skills from the virtual classroom to ECTs live classrooms is currently in progress but we present usage data of the system by teachers from three high-poverty schools in a North-Eastern State (Section 5.5).

3. Related Research and Requirements for Virtual Training Environment for Teachers

Although existing VTE-Ts have proven effective for some aspects of training, there does not seem to exist a comprehensive set of requirements to guide the development of, and improve research on, VTE-Ts.

As summarized in Table 1, one of our contributions is (1) to put forth a set of initial requirements that need to be considered before and during the development of a VTE-T based on the project given resources, and (2) to document how the most advanced VTE-Ts have addressed these requirements. To that end, we conducted a survey of the literature by rendering explicit the implicit chosen requirements of existing VTE-Ts research projects. We compiled our proposed set of requirements based on our analysis of existing VTE-Ts, on the initial set of requirements that education experts on our team had requested, and on requirements that emerged during our human-centered 4-year long lifecycle that led to our final software, IVT-T 4.3.

Our results are a set of six main requirements for VTE-Ts, shown in Table 1, that we consider desirable (if not necessary) for the development of future VTE-Ts: namely: *behavioral fidelity*, *environment fidelity*, *instructional design*, *autonomy*, *interactivity*, and *scalability*.

3.1. Behavioral Fidelity

Behavioral fidelity refers to the realism and consistency of the virtual human behaviors in the VTE, does a virtual aggressive 6th grader behave as a real aggressive 6th grader would in a real classroom, or, does an off-task virtual 1st grader behave as a real off-task 1st grader would? In the

Table 1
Comparison of VTE-Ts based on the proposed requirements

	Virtual Training Systems for Teachers	ClassSim Kervin et al. (2006)	SimSchool Collum et al. (2020)	Virtual PREX Dalgarno et al. (2016)	TeachLive Dieker et al. (2015)	3B Lugrin et al. (2019)	SimInClass Kelleci and Aksoy (2020)	IVT-T Delamarre et al. (2017)
								
Behavioral fidelity	Education experts' content	1 Scenario	Student model	12 Scenarios	Set of instructions	Set of instructions	16 Scenarios	9 Scenarios
Environment fidelity	Affect Graphics	Still facial expressions 2D Low Fidelity	COVE Model 2D Low Fidelity	None 3D	Facial expressions 3D	None 3D	None 3D	None 3D
	Student/Class Teacher representation	- None	1 to 18 None	10 8 avatars (4 females)	5 None	24 1 avatar (1 white male)	- None	15 None
	Animations	None	Still postures	~20 animations	Live motion capture	~25 animations	-	~80 animations
	Audio	No	No	No	Instructors' voice	20 recordings	No	~450 children recordings
Instructional design	Adaptability to user	None	None	None	Instructor's initiative	Instructor's initiative	16 difficulty levels	3 difficulty levels
	Reflection	Free-thinking space	None	None	None	None	None	Guided by questions
	Feedback	None	Student and classroom learning graph	Videos of recorded sessions	Instructor's feedback and recorded sessions	Instructor's feedback during and after session	Student and classroom learning graph	Autonomous feedback
Autonomy	System autonomy	Autonomous	Autonomous	Controlled by users	Controlled by instructor	Controlled by instructor	Autonomous	Autonomous
	Human resources	None	None	As many as student avatars	Trained instructor	Instructor	None	None
Interactivity	Technological requirements	Desktop and online access	Desktop and online access	Desktop and online access	Large screen and motion tracking device	HMD and motion tracking device	Desktop, HMD, Mobile and online access	Desktop and online access
	Type of interactions	Menu choices	Menu choices	Text chat menu choices	Voice and gesture	Voice and gesture	Menu choices	Menu choices
	Usability and UX	-	Training required	Cheat sheet required	Training for instructor	Intuitive UI for instructor	Issues on mobile	Easy to learn easy to use
Scalability		No	Yes	Yes	Yes	Yes	Yes	Yes

virtual agent community, behavioral fidelity when combined with graphical fidelity, is often referred to as believability, or the ability of the virtual entity to provide the "illusion of life" and suspend disbelief in users [Bates et al. \(1994\)](#).

Experiential learning theory [Kolb and Kolb \(2009\)](#) hypothesizes proposes that mastering skill developments is achieved through modeling and concrete learning experiences in realistic settings. Hence realistic scenarios depicting the virtual behavior for the entities in VTEs are essential to provide concrete and realistic situations, and to give users a sense of presence (i.e. the sense of "being there" in the VTE which has been positively associated with learning) [Mikropoulos and Natsis \(2011\)](#). Relevant scenarios play a major role for learners' motivation, and others have also pointed to behavioral fidelity as necessary for an efficient transfer of learning [Bossard et al. \(2008\)](#); [Dalgarno and Lee \(2009\)](#).

Three main approaches have been taken to generate virtual students' behaviors:

- scenario-based: creating all scenarios for each behavior at any given time that are then computerized and automatically controlled by the VTE-T,
- WOz-based: relying on a human instructor to control the virtual students' behaviors without the users' knowledge, according to a set of instructions provided in advance, and
- model-based: creating a student behavioral computational model that controls the virtual students' behaviors based on the current values of the model parameters during the simulation.

3.1.1. Scenarios-based approach

Education experts can create scenarios by mapping out exactly how the virtual students should act and react to teacher trainees' input during the virtual training simulations. These are then in turn programmed into the VTE-T to automatically control the virtual students. This approach was adopted for *ClassSim* [Kervin et al. \(2006\)](#) to generate a total of one scenario with 500 nodes.

[Dalgarno et al. \(2016\)](#) *VirtualPREX* uses *Second Life* and [Dalgarno et al. \(2016\)](#) is based on twelve role-play scenarios. Teacher trainees teams up to play the different roles of the scenarios with avatars. A trainee is attributed the role of the teacher while others play the students. The student roles were divided into on-task active, on-task passive, off-task active and off-task passive behaviors. Therefore to train one teacher, others are required to play the students.

For *IVT-T*, our education experts created 9 scenarios (3 scenarios for an off-task 1st grader, 3 scenarios for an off-task 6th grader, and 3 scenarios for an aggressive 6th grader). We computerized these scenarios as described in [Section 4.4](#), and evaluated them in terms of behavioral fidelity as described in [Section 5.1](#). For each play of a scenario, teacher trainees will be faced with making 4 to 8 decisions. Overall, the scenarios contain around 50 different decisions.

3.1.2. WOz tele-operation approach

A second approach to create realistic behaviors for the virtual students is to rely on experienced instructors to control the virtual students without the users knowing about it. This set up is known as a Wizard of Oz (WOz) in which a human uses controls to tele-operate the actions of the system, behind the users' view so that users have no idea the system they are interacting with is controlled by a human. Since this setup requires a human expert to be available when trainees need to use the VTE-T, it is related to the *autonomy* requirement, which we discuss later.

Whereas WOz setup has the advantage of giving freedom of users' input and adapting the system response to these inputs (i.e. trainees can try any strategy to address the behavior), depending upon the size and complexity of the instructions, this setup can create a significant cognitive load for the WOz instructor to manage.

In *TeachLive* [Dieker et al. \(2015\)](#) for example, the WOz observes a

teacher trainees' non-verbal behaviors and utterances, and takes control of one virtual student at a time (out of five students in total) to react to the trainee, while the remaining four virtual students automatically express passive behaviors. When taking control of one student, the WOz displays the corresponding posture, vocal utterances and reactions according to a student persona, provided to the WOz ahead of the training session. Whereas *TeachLive* usability has been evaluate from the trainee perspective, no information was provided about the usability of the system from WOz operator's perspective.

Breaking Bad Behaviors (3B) [Lugrin et al. \(2019\)](#) also makes use of a WOz to control the virtual students' behaviors. During 3D simulations, the WOz can adjust the level of disruption, [de/]activate a bad behavior out of six different behavior types, or twenty different dialogs by selecting a student and attributing a behavior to this student through a UI. In 3B, the WOz operator also needs to control the virtual environment points of view (overall situation, student's behaviors and reactions of the teachers), the camera control (front view of the class, back view, teachers' view point), and the feedback board to post feedback for the teacher. A usability evaluation for the WOz UI was conducted and results indicated that it was easy for the Woz to control the classroom.

3.1.3. Model-based approach

Two VTE-Ts, *SimSchool* [Gibson \(2011\)](#) and *SimInClass* [Kelleci and Aksoy \(2020\)](#) resorted to student models. *SimSchool* which supports the training of pre-service teachers for students with physical disabilities, created personality models for virtual student behaviors representation [Collum et al. \(2020\)](#). Each student is represented by a set of variables representing students' personality traits, academic level and physical-perceptual aspects, each containing 20 possible values. *SimSchool* can generate 20⁹ quantitatively different students. By altering the 3 physical-perceptual variables (e.g. by setting the vision, hearing and kinaesthesia variable(s), a student can portray physical disabilities so that some types of teacher-student interaction will not be effective (e.g., verbally addressing a deaf student will most likely fail).

The student in *SimInClass* are controlled by a Belief-Desire-Intention model based on social learning theory [Köknar \(2015\)](#).

3.1.4. Validation

Regarding how each VTE-T validated the behavioral realism of their virtual students in the scenarios, we found that not all VTE-Ts provided evaluation results. No evaluation for the behavioral realism of scenarios, or set of instructions provided to the WOz to generate these scenarios, were found for *SimSchool*, nor for *TeachLive* VTEs. A study of *SimSchool* conducted with 22 student teachers observed that some users disliked the lack of realism from virtual students' responses [Badiie and Kaufman \(2015\)](#), which could indicate that the model-based approach needs to be refined.

3B conducted an evaluation of the simulation effects and found that their eleven subjects rated the simulation effects equal or higher than scale average using teach-live questionnaire [Hayes et al. \(2013a\)](#), but that the subjects were most unsatisfied by the similarity between virtual and real students' behaviors, and in particular the low-arousal behavior (e.g. sleeping).

VirtualPREX [Dalgarno et al. \(2016\)](#) scenarios were not evaluated by education experts either but they were refined based on feedback from education student participants during a pilot-study. Validating the realism of the classroom interactions can avoid teachers to disregard the system [Badiie and Kaufman \(2015\)](#) and prevent a potential break in the sense of presence [Dalgarno and Lee \(2009\)](#); [Moskaliuk et al. \(2013\)](#). The example of *VirtualPREX* also shows that users can contribute to improve the content of the system.

Given that one of our main goals was to ensure high behavioral fidelity which research indicates is necessary for efficient transfer of learning [Bossard et al. \(2008\)](#); [Dalgarno and Lee \(2009\)](#), *IVT-T* behavioral fidelity was evaluated multiple times: the first behavioral fidelity evaluation (discussed in details in [Section 5.1](#)) was conducted with a

board of six retired teachers who, iteratively, rated four prototype versions of each scenarios. The evaluation was conducted on multiple dimensions, including the *logic and realism of students' behavior, actions, and dialogue according to their age* (1st or 6th grader) and *presenting problem* (inattention/hyperactivity or aggression/noncompliance), as well as the *teacher trainee response options*, and the *storyline engagement*.

We therefore suggest that VTE-Ts consider ensuring some level of behavioral fidelity, and that metrics such as the ones used for validating IVT-T, or the ones found in the teach-live questionnaire, be used for validating VTE-T behavioral fidelity.

3.2. Environment Fidelity

The environment fidelity requirement is a non-functional requirement describing how close to reality the look and sounds of the virtual environment is. This requirement includes the graphical realism of the environment, the animations, and the audio components. In the four-dimensional framework proposed by de Freitas et al. [De Freitas et al. \(2010\)](#), the *Representation* dimension includes the concept of fidelity. This concurs with the learning model of 3D VTE proposed by Dalgarno et al. [Dalgarno and Lee \(2009\)](#) in which representational fidelity in a central characteristics to generate users' sense of presence, co-presence (i.e. the sense of "being there with someone") within the VTE [Bailenson et al. \(2005\)](#); [Lee \(2004\)](#). Dalgarno et al. identify the textures, lightning, 3D models, frame per seconds, smooth view changes, spatial audio, and user representation as factors of the representational fidelity characteristic.

3.2.1. Graphics

There is an ongoing debate towards the realism of graphics in VTEs between researchers. Some posit that graphical fidelity could be detrimental for learning [Brenton et al. \(2005\)](#); [Wages et al. \(2004\)](#) while others argue that it is necessary for an efficient transfer of learning [Bossard et al. \(2008\)](#); [Dalgarno and Lee \(2009\)](#) and technology adoption [Ludwick and Doucette \(2009\)](#); [Whyte et al. \(2015\)](#). A study on anxiety during job interviews revealed that high fidelity graphics engendered higher levels of anxiety as well as a higher sense of presence [Kwon et al. \(2013\)](#). Finally, a third approach advocates for a graphical realism adapted to the type of learning [McLaughlin et al. \(2010\)](#). Additionally, [Bailenson et al. \(2005\)](#) showed that a mismatch between graphical realism and behavior realism have a negative impact on co-presence, indicating that the behavior fidelity requirement and the realism of graphics need to be both considered for the development of VTE-Ts.

In IVT-T, based on our readings of the literature which currently tends to support for higher (albeit not photo-real) graphics for VTEs, we aimed at 3D high realism and quality for IVT-T's graphics. We relied on iterative feedback from education experts to reach their desired level of realism. The *VirtualPREX* [Dalgarno et al. \(2016\)](#) classroom simulator did not perform an evaluation of their graphics but collected feedback from users. Feedback on *VirtualPREX* classrooms included the *proportionality of the furniture* (tables and chairs) while feedback on virtual students included clothes (school uniform) and faces. Some of the *VirtualPREX* avatars had adult faces on kid's bodies. Depending on the context of the VTE, *age appropriateness* of the virtual humans is a great addition to the evaluation of graphical realism.

3.2.2. Number of students per classroom

The number of students per classroom (reflecting real classroom settings) is also an important factor when considering the environment fidelity requirement as class size is a factor of teachers' stress levels [Lugrin et al. \(2016\)](#). Some existing systems represent their classroom with a low number of students (5 students for *TeachLive*, 10 students for *VirtualPREX*) whereas others, such as IVT-T, included a more realistic number of student to better reflect classroom sizes [Collum et al. \(2020\)](#); [Delamarre et al. \(2017\)](#); [Lugrin et al. \(2016\)](#).

3.2.3. User's representation

Graphically representing users in the environment can support the feeling of co-presence which in turn benefits social interactions [Schroeder and Axelsson \(2006\)](#). Of the existing systems, only *VirtualPREX* [Dalgarno et al. \(2016\)](#) and *3B* [Lugrin et al. \(2019\)](#) use avatars to represent the teacher. *VirtualPREX* created 8 teacher avatars (4 males and 4 females) for teacher trainees to choose from. Only 1 white male avatar seems to be available for the *3B* VTE-T. Studies have shown that users had more affinity and acceptance to a users representation that looked like them [Hayes and Johnson \(2019\)](#); [Lugrin et al. \(2015\)](#), but underlined the need for more research on more diverse user representation and their potential impact on user engagement, satisfaction and learning outcomes [Hayes and Johnson \(2019\)](#).

3.2.4. Audio

[Dalgarno et al. \(2009\)](#), in their learning model of 3D VTEs, defend spatial audio as a factor for the representational fidelity. In a study evaluating the user experience of *SimInClass* [Kelleci and Aksoy \(2020\)](#), some teacher participants were getting bored because of the lack of sounds. Teachers suggested to add sounds for a more realistic experience. Only 2 VTE-Ts resorted to audios. In *TeachLive* [Dieker et al. \(2015\)](#), the instructor WOz modulates his/her voice to impersonate the virtual students [Nagendran et al. \(2014\)](#). In *3B* [Lugrin et al. \(2019\)](#), the instructor can choose between 20 simple or advanced utterance recordings. However, [Dalgarno et al.](#) also defend for the "*Consistency of object behaviours*", which underlines some limitations of current approaches. Can an adult instructor realistically self-modulate his/her voice to impersonate 5 different middle school student? Similarly, [Lugrin et al.](#) do not specify if voice differences such as gender are considered within the 20 recordings used in *3B*.

IVT-T integrates three types of sounds: Classroom ambient sounds that plays in loop to give the feeling auditory feeling of being in a classroom; disruptive sounds that plays a specific times in the scenarios and aim at adding distraction to the teacher user and students' utterances that were recorded with children of matching ages to the virtual students. The realism of the students recording were validated by teachers during our user studies.

3.2.5. Validation

Representational fidelity is a distinguishing feature of teaching simulators that supports transfer of learning from the virtual classroom to the real classroom and predicts how quickly users adopt new technologies [Ludwick and Doucette \(2009\)](#); [Shernoff et al. \(2018\)](#); [Whyte et al. \(2015\)](#). Whereas some studies have looked into the evaluation of fidelity of virtual environment for driving simulations [Debattista et al. \(2017\)](#), urban planning [Drettakis et al. \(2007\)](#) and size perception [Geuss et al. \(2010\)](#), these metrics are not relevant for VTE-Ts. Therefore, we were interested in evaluating core elements of instruction that could potentially maximize learning and transfer. This included assessing the extent to which the main characters and non disruptive students were realistic looking, the classroom appearance was lifelike and believable, and whether we were immersing teachers in training scenarios that closely resembled their interactions with disruptive students. A board of experienced teachers was recruited to evaluate IVT-T's graphics realism (5.2). Additionally, qualitative feedback collected by the board of experienced teachers revealed other important aspects to consider for the virtual classrooms such as furniture proportions and lighting and for the virtual students such as body proportions skin tones, age appropriateness.

3.3. Instructional Design

The instructional design requirement aims at optimizing trainees' learning. This requirement concurs with the "*pedagogy*" dimension of the four-dimensional framework proposed by de Freitas et al. [De Freitas et al. \(2010\)](#) which considers learning and teaching models supporting

the learning included within the VTE. Even though instructional design does not appear in Dalgarno et al. affordances of 3D VTEs, Dalgarno et al. raised the question of how to integrate and adapt instructional elements for 3D VTEs [Dalgarno and Lee \(2009\)](#).

For our instructional design requirement we consider three main aspects: adaptability to user, users' reflection and experts' feedback.

3.3.1. Adaptability

The VTE adaptability to the user influences how the system can regulate the difficulty of a simulation to challenge the user while avoiding frustration. In their framework, Nadolski et al. argue that adaptation to the learner results in deeper and more meaningful learning [Nadolski et al. \(2012\)](#).

In the game design domain, researchers argue that challenge is a characteristic intrinsic to good video games [Malone \(1980\)](#). Goal achievement must be uncertain in order to keep players entertained. If the game is too easy, players are more likely to get bored and disengage and conversely, a game that is too difficult will generate frustration. The same can be applied to VTEs [Nadolski et al. \(2012\)](#). Therefore, in order to provide an effective training experience, the difficulty of the simulation needs to adapt to the user's current expertise.

In VTE-Ts relying on a WOZ tele-operated approach such as *TeachLive* [Dieker et al. \(2015\)](#) and *3B* [Lugrin et al. \(2019\)](#), the simulation difficulty can be dynamically adapted to the trainee by the instructor controlling the system. Ultimately, the difficulty of the simulation will be decided by the teaching style and skills of the instructor.

For VTE-T using model-based approaches (e.g. *ClassSim* [Christensen et al. \(2011\)](#)), the difficulty of the simulation is determined by the parameters with which each virtual student has been initialized and the range of differences between students. In *ClassSim*, virtual students initialized with different values necessitate different strategies to start learning. No details are provided on whether the students' parameters can be modified at run time to dynamically adapt the difficulty to the learner's expertise level.

Finally, for systems using a scenario-based approach, since interactions with the virtual students are pre-scripted, it is not possible to dynamically adapt the simulation, however other techniques exist to maintain users' interest such as level-up systems [Jemmali et al. \(2018\)](#). *SimInClass* [Kelleci and Aksoy \(2020\)](#) created 16 difficulty levels. The increase in difficulty between levels was represented by the number of students and the frequency of unwanted behaviors. In IVT-T, we took a similar approach. A total of nine scenarios with three difficulty levels were developed. The more complex scenarios can be unlocked by completing objectives in the simpler scenarios. The IVT-T's adaptation to the trainee happens between scenarios, where each trainee can train at their own pace, with scenarios matching their current behavior management skills level.

3.3.2. Reflection

Traditional professional development and on-the-job-training provide limited opportunities for ECTs to reflect and problem solve different CBM approaches. Previous research has shown that learning and transfer increase when reflection is integrated into the instruction [Merrill \(2009\)](#). Giving teachers the opportunity to reflect within the simulator is thus crucial to design an efficient training system for teachers. Integrating a reflection space can also provide great qualitative feedback on the use of the system by teachers [Kervin et al. \(2006\)](#). Previously, only *ClassSim* [Kervin et al. \(2006\)](#) integrated a reflection space for pre-service teachers ([Table 1](#)). Authors noticed that pre-service teachers used this feature significantly, and some pre-service teachers would even copy-and-paste parts of the educational resources provided into their thinking space.

We therefore leveraged that knowledge for IVT-T development, and included a specific feature for trainees to enter their reflections. IVT-T training sessions are therefore composed of 4 phases: Practice, Replay, Reflect, Feedback.

IVT-T was developed using three critical pedagogical design elements that are empirically linked to learning and skill acquisition [Dede \(2009\)](#); [Killion \(2013\)](#); [Salas et al. \(2012\)](#):

- **Phase 1 Practice** (i.e., teachers traversing vignettes and making decisions regarding how to respond to the character) was embedded into the system to allow teachers extensive opportunities to experiment with how to respond to disruptive behaviors in a safe environment. This approach to training teachers has ethical advantages to practicing with live students and promotes learning that has application to later performance in the classroom.
- **Phase 2 Replay** provides an opportunity to teachers to replay their entire interactive session as a reminder.
- **Phase 3 Reflection** (i.e., asking teachers open-ended questions regarding the choices they made and how those choices impacted the student) was used to promote critical thinking and new learning. Reflection was designed to help teachers re-examine their interactions with characters and how their responses contributed to increasing and decreasing problem behavior.
- **Phase 4 Feedback** (i.e., sharing information about the effectiveness of the teacher response to the character) was designed to increase teacher acquisition of knowledge and skill in preventing and responding effectively to disruptive behavior. It was hypothesized that the learning curve would be accelerated as teachers adapted their responses to disruptive behaviors based on immediate feedback regarding their performance.

To assist ECTs in their reflections and help them remember specific decisions, the system provides visual cues to the trainee from specific decision points in the simulation (screenshots of the simulation when Jordan swore and kicked his desk), accompanied with questions such as "Explain why the student reaction surprised you" or "Explain why you wished you have made a different choice". By providing reflection questions IVT-T aims to direct ECTs reflection on why the disruptive behavior occurred, the relative effectiveness of the strategy they selected and how they plan to improve in the next IVT-T simulation.

3.3.3. Feedback

During traditional training, teachers have few opportunities to practice while receiving feedback [Denton and Hasbrouck \(2009\)](#); [Shernoff et al. \(2015\)](#). Providing feedback is a necessary requirement for the creation of an effective teacher's training system and most existing system provide experts' feedback (see [Table 1](#)). Instructional design researchers indicate that practicing without explicit feedback does not result in strategy retention or transfer of learning [Richey et al. \(2011\)](#); [Tracey et al. \(2014\)](#).

Feedback can be provided at run-time during the interaction with the simulated students, or after the simulation interaction. [Lugrin et al. \(2019\)](#) observed that giving audio feedback cues during the simulation did not affect the feeling of presence or the suspension of disbelief of users. Generally, classroom simulators provide feedback after the simulation as quantitative or qualitative data. For instance, *SimSchool* [Collum et al. \(2020\)](#) generates graphical representation of the evolution of virtual students' learning over time, as well as an overall graphical representation of classroom teaching effectiveness. In *VirtualPREX* [Dalgarno et al. \(2016\)](#), video of recorded sessions are used as support to give feedback to teacher trainee.

The goal of the feedback in IVT-T is to activate ECTs' prior knowledge of CBM strategies, maintain their engagement with the system, help them acknowledge how they performed, and adapt their approaches to new situations to improve their performances [Dieker et al. \(2015\)](#); [Lugrin et al. \(2019\)](#); [Schunk \(2008\)](#). Hence, following the reflection phase, IVT-T provides ECTs with feedback on the choices they made to address the disruptive student behavior, that was designed by education experts. IVT-T experts' feedback is quantitative (~250 quantitative feedback entries per vignette) and qualitative (~200

qualitative feedback entries per vignette).

3.3.4. Validation

The evaluation of the instructional design requirement is directly linked with the efficiency of the training provided by the system. This can be observed by monitoring the performance of ECTs being trained with the VTE-T compared to a control group following a traditional CBM training (which is our fourth research objective shown in Section 2.3).

Evaluating the efficiency of the training provided by the system requires summative evaluations of the final training system that measure the training impact on real teachers CBM's skill transfer and on their real students' learning outcome, which are that conducted over time by domain experts in real schools, with consent from school superintendents, principals, teachers, and parents' consent.

Before that summative evaluation becomes possible with the final training system, it is possible to observe how ECTs perceive and use the different aspect of the instructional design with formative evaluations. Feedback on the system can be collected using interviews or focus groups.

For example, access to IVT-T was provided to a group of teachers from different schools to observe their interactions with the system. Through focus groups, different elements of IVT-T's instructional design were thus evaluated, discussed, and revised accordingly.

3.4. Autonomy

Autonomy is the extent to which a system can be used independently by a user without requiring another human. This requirement determines whether trainees can train their CBM skills independently, or whether they need to wait for an instructor and/or human operator to run the system for them anytime they wish to practice. Technological devices required are also included in the autonomy requirement as these directly impact ECTs' ability to train with the system.

3.4.1. Human resources

Existing classroom simulators using role playing or WOz setups (see Sec. 3.1.2) increase the man power needed to run the system so that trainees can practice. In the case of *VirtualPREX* Dalgarno et al. (2016), to train one teacher, the system requires ten human actors to play (i.e. teleoperate) the ten virtual students present in the classroom, whereas *TeachLive* Dieker et al. (2015) and *3B Lugin* et al. (2019) cannot provide training without the interventions of trained instructors controlling the system.

The use of human actors can make it difficult to expose teachers to identical situations. Moreover, recreating similar scenarios can become a challenge for the instructor controlling the system. Requiring personnel to run the system greatly reduces how much practice a teacher can receive with the VTE-T and increases the cost of training Dawson and Lignugaris/Kraft (2017).

Populating the virtual classroom with autonomous virtual students can overcome this constraint. However, as shown by the study led by Badiee et al. Badiee and Kaufman (2015), an autonomous system using a model of behavior, such as *SimSchool* can result in users questioning the realism of the interactions and of the situations presented, unless the model is thoroughly validated.

Using the scenario-based approach (see Sec. 3.1.1, our IVT-T system is completely autonomous: ECTs can practice, reflect and receive feedback without having to rely on any human actor. Moreover, IVT-T uses scenarios which were validated in terms of realism by experienced teachers (Section 5.1). Therefore, IVT-T provides a training platform with which ECTs can practice autonomously on realistic classroom situations.

3.4.2. Technological requirements

The autonomy requirement also needs to address when the system can be used, where the system will be set up, and what technology is

required for the system to run in order to ensure that the training environment meets the user's needs.

Of the presented VTE-Ts, *SimInClass* Kelleci and Aksoy (2020) is the only one offering the use of the classroom simulator on mobile devices. However, during a user experience study of the system, teachers had issues interacting with the mobile version of *SimInClass*. Some user interface element would not be displayed correctly thus hindering the system's usability. Additionally, some teachers indicated that they preferred the Desktop version because it provided a larger screen.

Existing systems such as *TeachLive* Dieker et al. (2015) and *3B Lugin* et al. (2019) require hardware equipment (Head-Mounted Device (HMD) and motion tracking device) that necessitates expertise and an adapted motion capture laboratory space that may not be available in end-users' environment.

A system that can adapt to its users' habits and accessible from any location (workplace, home, etc.) allows teachers to choose when they practice and can encourage teachers to practice their behavior management skills as much as possible. Location and access to training are two aspects considered for the "Context" dimension proposed by de Freitas et al. De Freitas et al. (2010).

IVT-T is available online through a website and only requires a laptop or desktop computer to run. By downloading the IVT-T applications, ECTs have access to different realistic scenarios and associated feedback as well as pedagogical resources.

Nevertheless, with the development of immersive virtual reality technologies such as HMDs, more studies are needed to evaluate the impact of technology using virtual humans on the transfer of learning. Recent work by Ochs et al. Ochs et al. (2018) on the training of communication skill for medical experts indicates that more immersive technologies such as HMDs or Cave Automatic Virtual Environments (CAVEs) result in more presence and co-presence than a desktop setup and potentially more transfer of learning. However, given the limited number of participants (n=22, 11 of them being actual doctors), Ochs et al. acknowledged that no conclusion could be drawn for the use of immersive VR technology in a social context and encouraged to conduct larger experiments to confirm their results. However, Lugin et al. Lugin et al. (2016), observed that pre-service teachers responded positively to the use of immersive virtual reality for the training of classroom management skills.

3.4.3. Validation

The main concern of the autonomy requirement is to give ECTs access to practice as much as possible. A system that allows ECTs, whose schedule can be busy, to choose when and where they want to practice can achieve that. The number of other human actors required for ECTs to practice is also a factor in the evaluation of the autonomy requirement.

Systems requiring personnel, facility, and equipment impose constraints on how much ECTs can practice. For instance, in a study on the use of praise with *TeachLive*, some data point were missing because of teacher trainees absence. Trainees had to go to a different location than their school to practice Dawson and Lignugaris/Kraft (2017). This example illustrates limitations of systems requiring ECTs to go out of their way to access the training system.

3.5. Interactivity

As mentioned earlier, experiential learning theory Kolb and Kolb (2009) hypothesizes that mastering skills (e.g. proactive monitoring, effective redirection) is achieved through concrete and physical learning opportunities in realistic learning situations. The IVT-T approach is to provide experience with dealing with disruptive behaviors in a classroom context, which can be addressed by ECTs deploying proactive monitoring strategies, and using effective redirection approaches. By making decisions and choices in the virtual classroom and by visualizing their decision consequences on the virtual students behaviors, IVT-T aims to support reflection on which approach is working and which

one is not for a given context Kolb and Kolb (2009); Lindsey and Berger (2009).

Additionally, interactivity appears as a critical component in many development frameworks for VTEs De Freitas and Oliver (2006); Moskaliuk et al. (2013); Nadolski et al. (2012). Dalgarno et al. Dalgarno and Lee (2009) identify interactivity, including embodied actions and embodied verbal and non-verbal communication, as a main characteristics of 3D VTEs to generate a sense of presence, co-presence in the user.

Our proposed interactivity requirement includes three components: the type of interactions (i.e. how teacher trainees can act and communicate within the VTE), usability, and UX. Systems that are easy to learn, efficient to use, and pleasant to interact with, avoid cognitive overload that can negatively affect users' learning experience and willingness to engage with the system Hartson and Pyla (2019); Nielsen (1994); Roldán-Álvarez et al. (2016).

3.5.1. Interaction modalities

From the presented VTE-Ts, we distinguish 3 types of interaction modalities: menu choices, live typed text chat, and verbal and non-verbal communication.

In ClassSim Kervin et al. (2006), depending on the context of the situation, different menu options are presented to the trainees such as reprimanding or ignoring students who speak without raising their hand, intervening or not when students are pushing one another to enter the classroom, or encouraging reserved students to participate in classroom activities.

SimSchool Collum et al. (2020) interactions are divided between two types of action: (1) attributing or adjusting tasks (asking to recite a poem or to work alone at one desk), or (2) addressing a virtual student with an utterance. Users select from multiple option menu for each of these two types

In VirtualPREX Dalgarno et al. (2016), teachers playing different roles communicate through Second Life live typed text chat, and can select animations to be played by their avatars with a menu. Gregory et al. observed, however, that some users needed cheat sheets to remember how to control avatars via Second Life user interface. TeachLive Dieker et al. (2015) and 3B Lugrin et al. (2019), being controlled by a WOz that observe users' actions and trigger a system reaction, allow users to interact with verbal and non-verbal communications with the system.

In IVT-T, we chose menus for enabling teachers to read each of the ABC strategies so that they have time to identify praises; we furthermore use student's verbal communication to be displayed in a bubble as well as spoken by children's recorded voices for teachers to be able to hear and read the students utterances.

3.5.2. Usability and UX

Usability is a measure of the efficiency and satisfaction of user interactions with a system and UX is concerned with the overall experience of a user interacting with a system.

Validating usability and UX throughout the design lifecycle development process is necessary to identify usability problems early and helps guide the overall development of a user-friendly system Hartson and Pyla (2019); Chellali et al. (2016). Moreover, trying to correct usability issues later in the development process can be very costly and time consuming Hartson and Pyla (2019); Bowman et al. (2002).

SimSchool Collum et al. (2020) and *VirtualPREX* Dalgarno et al. (2016) faced the issue that users were having difficulties interacting with the system, so users resorted to cheat sheets to the side of the simulator to remember commands. Studies evaluating *SimSchool* showed that it took approximately one hour for teachers to familiarize themselves with the UI Rayner and Fluck (2014). Additionally, teachers disliked *SimSchool*'s UI because they had difficulties navigating the options Badiee and Kaufman (2015). These studies show that UI designs can trigger teachers' frustration with the system itself, which is not conducive for learning the necessary skills. It is therefore recommended

that VTEs designers adopt the human-centered design lifecycle that uses iterative cycles of design, prototype, evaluate, to produce final UIs that are usable and pleasant to interact with. For IVT-T design and development, we adopted rapid prototyping of UI designs via storyboards which we evaluated and revised, before programming more extensive UI features, which were also in turn evaluated and refined before leading to the final UI for IVT-T 4.0 (see Fig. 2).

3.5.3. Interactivity Validation

To evaluate usability and UX, we suggest a mixed-method approach, one which includes the collection of both quantitative data (e.g. standardized rating scales and questionnaires) and qualitative data (e.g. observations of user performance and semi-structured interviews). Using both quantitative and qualitative data helps provide a more thorough and complete understanding of how a system is perceived by users, compared to using quantitative data only or qualitative data only Creswell and Plano Clark (2011). Whereas quantitative data provide good overview of the usability and UX problems, qualitative data can provide insights about *the reasons* for the usability and UX problems (which cannot be inferred from quantitative data), and help designers find solutions.

Quantitative data can provide insights into the usability of the system and can be used to compare it to existing technologies. This data can provide indication on user attitudes toward the system in terms of ease of learning, ease of use and technology adoption Ludwick and Doucette (2009).

To evaluate IVT-T, we used well accepted self-report questionnaires provide standardized measures to collect quantitative data, such as the System Usability Scale (SUS; Bangor et al. (2008); Brooke (1986)) or the Questionnaire for User Satisfaction (QUIS, Chin et al. (1988)). QUIS provides data on the overall usability of the system, screen design and layout, terminology, learning, and system capabilities. Moreover, we collected qualitative data to reveal usability problems and design flaws (which cannot be inferred from quantitative data). We used methods such as CTA, during which users explained orally what they were thinking as they perform tasks assigned by a facilitator/observer next to the user Cooke (2010); Jaspers (2009). CTAs can reveal specifically what issues participants are having with the UI as they work through tasks. CTAs also generate real-time feedback and emotional responses to the system, which are good indicators of UX. Additionally, semi-structured interviews were used to obtain a better understanding of user satisfaction with the graphics and/or instructional design elements Morgan (1996).

3.6. Scalability

Scalability, in this context, refers to the system ability to integrate scenarios or situations without requiring major changes in the implementation. This requirement coincide with Lugrin et al.'s "extensibility" requirement for 3B Lugrin et al. (2019).

3.6.1. Scenario variability

In real classrooms, teachers face many different situations which vary depending on factors such as the context, the size of the classroom, the academic level, the type of disruptive behaviors, and the number of students. Presenting different training situations is desirable for VTE-Ts as it generates abstraction which benefits the transfer of learning Boscard et al. (2008). VTE-Ts implementing the scalability requirement have the ability to integrate additional classroom situations without having to change the current implementation of the system and thus can provide a variety of classroom situations to teachers.

For instance, *VirtualPREX* combines the Second Life platform with role-play scenarios. By using the text chat and the animations provided in Second Life, it is possible to generate different role-play scenarios that can be directly played in the virtual environment. No further implementation is needed to add new scenarios. Similarly, *TeachLive* and *3B*

can easily generate new classroom situations as they are controlled by instructors. Instructors can choose at anytime to integrate variation in the classroom situation the users is facing. *SimSchool*, using a model-based approach (see Sec. 3.1.3), can generate different reactions from the virtual students by changing the model parameters. For the *ClassSim* simulator, users interact with the system through a sequence of static *html* pages. To offer new content to their users, new static *html* pages must be created with their associated content (2D images, buttons).

IVT-T uses MASCARET Querrec et al. (2004) a meta-model that provides a description of a virtual environment by interpreting UML concepts and particularly UML activity diagrams. By translating scenarios into UML activity diagrams, IVT-T can play any scenario written by the education expert team without modifying the implementation of the system (this process is detailed in later sections). Using MASCARET, there is no limit to the number of students involved in a scenario, their behaviors, or even the size of the scenario itself, provided that the corresponding 3D models and 3D animations are available. To integrate a new scenario, IVT-T requires education experts to write down the scenario as a decision tree respecting specific node shapes for each actors. Hence IVT-T is highly scalable or highly extensible.

3.6.2. Scalability Validation

To evaluate the scalability of a system, one can count the number of modifications required in order to simulate a new classroom simulation. For instance, *ClassSim* requires to create a whole new set of 2D static screens to present new scenarios to their users (which includes generating the graphics for each screen and the links between each screens). On the other hand, the instructor controlling the 3D classroom, can attribute different actions to different students to generate new situations.

In IVT-T, because we translate scenarios from decision tree into UML activity diagrams, once new scenarios are written by domain experts, IVT-T can play the scenarios without modifying the implementation of the system (more details later).

3.7. Requirements Summary

In this section, we presented the requirements we claim beneficial for the development of VTEs for teacher training. These include: (1) *Behavioral Fidelity* to ensure that the behaviors and situation presented to users are similar to what teachers can face in reality; (2) *Environment Fidelity* have been shown to enhance the sense of presence (positively associated with learning Moskaliuk et al. (2013)) and can potentially improve learning Dalgarno and Lee (2009); (3) *Instructional Design* including feedback and users' reflection to support learning Merrill (2009); Richey et al. (2011); Tracey et al. (2014) as well as adaptability to users' level of expertise to maintain engagement with the system Nadolski et al. (2012); (4) *Autonomy* to ensure teachers an autonomous and easy access to the system for a better integration of the VTE in their schedule. (5) *Interactivity* which appears in all VTE development frameworks and is argued to generate a strong sense of presence, co-presence Dalgarno and Lee (2009); De Freitas et al. (2010); Moskaliuk et al. (2013); Nadolski et al. (2012); and (6) *Scalability* equip the VTE with the ability to present a variety of situation to users with few changes to the current implementation which support users' abstraction and thus the learning Bossard et al. (2008);

4. Interactive Virtual Training for Teachers (IVT-T)

This section describes first the goal of IVT-T and requirements that were provided by the education expert team. We then present our approach to build the IVT-T classroom simulator based on these requirements and on the requirements proposed in the previous section.

4.1. Overview

By visiting IVT-T website, ECTs have unlimited access to the IVT-T classroom simulator with which they can complete practice sessions. Practice sessions consist of completing exercises in the 3D virtual environment, watching replays of the simulation, reflecting on actions taken during the simulation, and receiving feedback about these actions. An IVT-T simulation is built from two main components:

- **Vignettes:** Vignettes are classroom scenarios designed to reflect real life situations experienced and created by our team of education experts. Vignettes map out the potential sequence of events based on ECTs' classroom management choices (some detrimental, some positive). Once realistic vignette scenarios are formatted for the IVT-T system, no other human input is required, therefore ECTs practice autonomously. To encode vignettes, we used the Multi-Agent Systems to simulate Collaborative, Adaptive and Realistic Environments for Training, (MASCARET) Querrec et al. (2004) which associates sequences of actions with virtual students within a 3D environment using UML concepts (A portion of a vignette is shown on the left of Fig. 6).
- **Virtual Environment:** IVT-T's virtual environment is composed of two 3D classrooms (1st grade and 6th grade) and 30 different 3D virtual students designed to reflect appropriate ages (15 1st graders and 15 6th graders). To ensure ECTs' immersion in the classrooms, efforts were concentrated on different physical arrangements reflecting the academic level, quality, and realism.

The simulator was built using Unity WebGL to ensure 24/7 online access from anywhere, and it can be run on any computer with a graphic card supporting 3D (25 frames per second with a middle class graphics cards such as the NVIDIA GeForce GT 750M). The loading time of the simulator in the browser can varies from 30 seconds to 2 minutes.

4.2. Architecture

The architecture of the IVT-T system is detailed in Fig. 3. It contains the four main components of an Intelligent Tutoring System (ITS): (1) the Domain Module, (2) the Learner Module, (3) the Pedagogical Module and (4) the User Interface (UI) Module Wenger (1987).

The Domain Module: This module contains the knowledge from the domain. In IVT-T knowledge is represented by vignettes, they are realistic scenarios of teacher-student interactions in classrooms. As the vignette unfolds, ECTs make decisions that will escalate or deescalate the disruptive behavior. Moreover, the domain module contains knowledge about scores and feedback for each decision in all vignettes, as well as the type of ending (positive, mixed or negative), i.e. the quantitative description of situations reached after the sequence of decision made by ECTs. The domain module includes reflection themes that contain different questions to guide the reflection of ECTs while they are practicing with IVT-T. The number of themes and questions can be expanded. Finally, the domain module also includes links to pedagogical resources.

The Pedagogical Module: It represents how the expert knowledge will be transmitted to the users. In IVT-T, education experts created practice sessions, each composed of four phases: Practice, Replay, Reflection, and Feedback.

1. **Practice:** ECTs make decisions while the vignette unfolds in the simulator. Depending on the level, they can choose to explore the range of possibilities proposed by the vignette. In level 1 for example, since no points are earned, they can choose to make the worst decision to see what happens. In level 2, the system starts keeping track of their scores, an incremental counter lets them know how long they took before making a decision. Therefore, ECTs can follow their progress as they complete practice sessions. Moreover, they need to

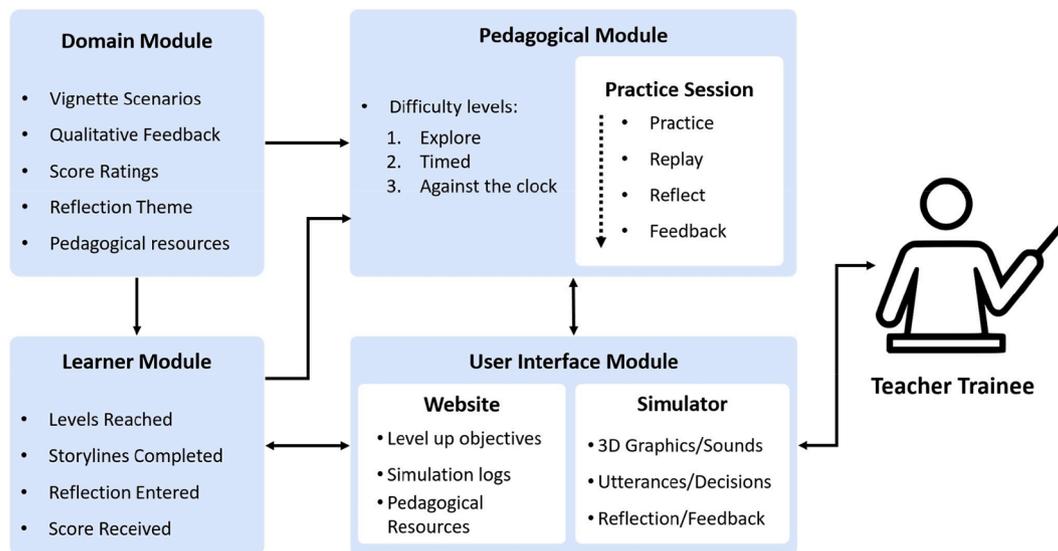


Fig. 3. Overview of the IVT-T system.

fulfill some conditions in order to obtain access to level 3. In level 3, their scores are also recorded, and they have a limited time to make a decision, before the system chooses for them the worst possible decision (if they have not made a selection).

2. **Replay:** After the practice, ECTs watch the replay of the practice. During this phase they can review the choices they made and observe again the students reactions.
3. **Reflect:** During the practice session, ECTs are encouraged to reflect on the decisions taken during the practice phase. Reflection themes from the domain modules help guide ECTs' reflection, which they can enter in their Reflection space which they can review anytime from their training log.
4. **Feedback:** In the feedback phase, ECTs view a sequence of screenshots picturing their decisions. On each screenshot, the quantitative feedback for all choices are displayed (score) as well as a qualitative feedback giving a textual comment on the decision and explains the important aspects to take into account.

To adapt the simulation difficulty to ECTs, a leveling up feature was implemented. ECTs must fulfill a certain number of conditions provided by the education experts such as *four storylines with a score higher than 80% must be completed or four different storylines have been completed*, before they can unlock a level providing more complex simulations, with more challenging type and intensities of behaviors.

The Learner Module: To assess the trainee's progress and learning, ECTs' knowledge is represented by their storylines, which are logs of the simulations they have performed in the system. Thus, we can track decisions made in each vignette and attribute scores to the overall storyline by counting the number of effective decisions and the difficulty levels reached. Finally, reflections made during the practice session are aggregated to the Learner module and accessible from the training log.

The UI Module: The IVT-T system has two main parts, the simulator and the website. In the simulator, the UI module informs ECTs with graphics and animations performed by the virtual students, spoken and written utterances, movements in the 3D environments, and icons. The reflection and the feedback phase of the practice sessions are also integrated in the simulator. In the website, ECTs can review their storylines, scores and reflections. The website also indicates which objectives must be completed to access the next difficulty level. Moreover, they also have access to other pedagogical content such as disruptive students' biography or an online course for traditional CBM lessons associated with concepts simulated in the environment.

As of today, IVT-T do not include an intelligent tutor. However, since

IVT-T uses an ITS architecture, future iterations of IVT-T could easily integrate a tutor that adapts instructions and feedback to each user.

4.3. IVT-T Graphics

This section presents the 3D classrooms and the virtual students with the behaviors they can display.

4.3.1. 3D Classrooms

Two 3D classrooms were designed for IVT-T, one 1st grade and one 6th grade. Significant effort was made to ensure realism of the classrooms to enhance ECTs' immersion, and a number of iterations of prototypes and feedback ensured their authenticity. Feedback was provided by six educators with many years of experience working in elementary schools. In real classrooms, teachers design and organize their classrooms to reflect specific age-groups.

Accordingly, to provide an immersive experience, classrooms also need a high level of realism. Thus, special features like wall decoration, table layout and furniture, were considered to enhance the verisimilitude according to the classroom grade.

In the 1st grade classroom (Fig. 4a) for example, a rocking chair and a carpet were placed in a corner, and the desks were organized into clusters. On the other hand, in the 6th grade classroom (Fig. 4b) computers were added, and, rather than the alphabet, scientific methods are displayed on the walls. Desks were organized uniquely, with all oriented in rows facing towards the board. The virtual classrooms, without the students, count 200,000 triangles.

The classrooms also incorporate different ambient sounds. The main ambient sound plays in a loop background noises of a working classroom. Additional ambient sounds occur only once, and vary from school announcements to police sirens passing by near the school. The vignette indicate when to play these additional ambient sounds.

4.3.2. 3D Virtual Students

Virtual students were developed using MakeHuman, an open source software able to create, rig and animate 3D characters. The features of the virtual students such as body shape, skin color and clothes, were customized to create unique virtual children. The number of triangles for the virtual students ranges from 11,000 to 34,000 (average is 22,000).

A total of 30 characters, 15 1st grader and 15 6th grader were designed. Each character has a unique skin color, hairstyle and body shape, illustrated by figure 5 showing the face of three 1st grade



(a) 1st grade, view from the back of the room



(b) 6th grade, view from the teacher's desk

Fig. 4. IVT-T Virtual Classrooms.



Fig. 5. IVT-T Virtual Students version 4 - 1st grade students.

students.

In order for virtual students to take actions in the virtual classrooms and thus to autonomously and realistically display the progression of the vignette scenarios, we resorted to 3D behavior animations.

First, the list of all possible behaviors was extracted from the vignettes provided by the education experts. Freely accessible online databases provided animations exhibiting common behaviors such as *Walking*. Behaviors that were specific to IVT-T vignettes (such as *Knock desk over* or *Middle finger to the class*), were recorded using 2 Kinects in stereo [Gao et al. \(2015\)](#) combined with a software linking depth maps with virtual humanoid skeleton (Iposoft).

The list of behaviors recorded is shown in [Table 2](#). Neutral classroom behaviors are the behaviors played by the non-disruptive students. On the other hand, off-tasks behaviors and aggressive behaviors are played by the disruptive students. First low-intensity behaviors are played to create a CBM situation, and the disruptive behaviors increase in intensity if the situation is not handled properly.

4.4. IVT-T Simulator

The main component of the simulator are the graphics (the classrooms models and the 3D virtual students), and vignette scenarios designed by education experts. The goal of the simulator is to play scenarios within the 3D environment with each student autonomously accomplishing their own actions and interacting with ECTs. This section describes the vignette component of IVT-T and how vignettes scenarios are translated into a 3D simulation.

4.4.1. Vignette

A total of 9 vignette scenarios were created in LucidChart. Different boxes (or nodes) are connected to each other with each node containing utterances and/or actions. By going through these sequences of utterances and actions, vignettes describe realistic scenarios of disruptive behaviors in a classroom context. A special type of node, decision nodes (yellow diamond-shaped box in [Fig. 6](#)), describes where ECTs need to

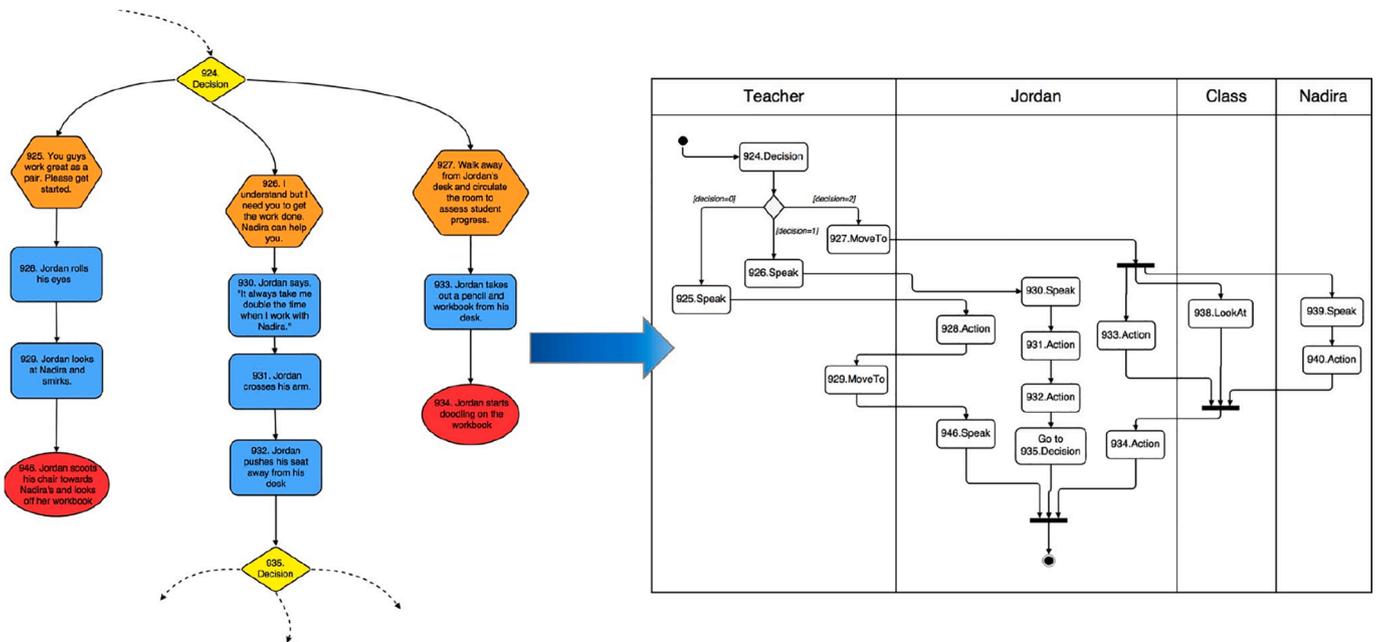


Fig. 6. Translation of a vignette sample to a UML Activity Diagram. Diamonds are interpreted as decision-merge nodes. Actions nodes for a given virtual student are attributed to its corresponding role in the activity diagram.

make a decision in order to advance in the scenario. IVT-T includes two main types of behaviors:

- **Disruptive behavior:** A disruptive virtual student can be off-task or aggressive. Off task and aggressive behavior examples are presented on the bottom of Table 2. The behaviors are displayed by the virtual student using animations and objects interaction (e.g. the *Knock desk over* behavior plays the knock desk over animation where the student swing his/her arms up while the desk is being tilted to the ground). The intensity of the disruptive behaviors displayed evolve based on the effectiveness of the selected choices. The disruptive student starts by displaying low intensity behaviors (specified by the vignette) and high intensity behaviors are being display if the situation escalates (ineffective CBM strategy).
- **Non disruptive behavior:** Non disruptive students are controlled by a finite state machine, looping through neutral classroom behaviors relevant to the context of the vignette (Top of Table 2). Neutral classroom behaviors are defined by behaviors specified in the vignette for the non-disruptive students. At the beginning of a

vignette, all non-disruptive students are initialized by playing an idle behaviors from the finite state machine which then starts looping in a random order so all students move differently while being idle. When an ECT makes an ineffective decision, the reaction to this decision is specified in the vignette for the disruptive students and sometimes for the non-disruptive students as well to represent the propagation of the disruptive student’s behaviors to other students. In this case, the finite state machine is interrupted to allow non-disruptive students display low-intensity off-task behaviors.

Each of the 9 IVT-T vignettes contains around 600 nodes with 70 different endings and around 80 different behaviors can be played by the students. To reach an end node (rounded red box in Fig. 6) ECTs go through three to eight decision nodes. The number of animation behaviors in IVT-T is above the ones of existing systems (between 20 and 25 animations for VirtualPREX Dalgarno et al. (2016) and 3B Lugin et al. (2019)).

4.4.2. From vignette to 3D simulation

By going through the vignettes, ECTs faced different classroom situations so they can intensively practice their behaviors management skills. In IVT-T, we consider a simulation as an ordered sequence of actions and choices realized by the different protagonists of the classroom according to the vignettes’ flow. During the progression of the simulation, disruptive behaviors are minimized as ECTs make effective choices. Conversely the student becomes more unruly and/or less willing to work if bad decisions are made.

Once a choice has been selected, the corresponding action is executed and the simulator starts performing the sequence that follows the action, with each student performing their own actions autonomously. ECTs can observe the consequences of their choice until another decision node or an end node is reached.

In order to create a simulation, vignettes are translated into a 3D interactive environment. However, vignettes are immense (600 nodes and around 50 decisions segments like the one illustrated in Fig. 6). Each node represents an utterance and/or behavior to be performed by one or many virtual students.

We used a multi-agent framework called MASCARET Querrec et al. (2004), a meta-model that provides a description of a virtual

Table 2

List of behaviors (neutral classroom behaviors, off-task behaviors, and aggressive behaviors) displayed by the virtual students in IVT-T.

Neutral classroom behaviors			
Idle sitting (4)	Writing Reading	Point to board	Open/Close book
standing Stand	Rummage desk	Point to paper	Flip pages Slide
up Sit down	Rummage	Twist on chair	book Sharpen
Walking Raise	backpack Take/put	Take/Put chair	pencil Work with
hand Scoot chair	on/in desk	Open/Close door	neighbor
		Walk with chair	
Off task behaviors		Aggressive behaviors	
<i>Low intensity</i>	<i>High intensity</i>	<i>Low intensity</i>	<i>High intensity</i>
Cover mouth Plop	Doodling Draw on	Tap pencil Play	Knock desk Push
on chair Shrug	hand Rocking	with phone	book Push chair
shoulders Lean	chair Roll pencil	Finger tapping	Slam book Slam
back Slouching	Spill paint Play	Wave hand	door Drum on
Elbow on desk	with paint Play		desk Foot kick
Head on desk	with Ipad Cross		Singing Middle
Head on arms	arms Listening to		finger Drop Ipad
Whisper to	music		Take out phone
neighbor			

environment by interpreting Unified Modeling Language (UML). MASCARET covers all aspects of a virtual environment semantic representation: domains ontology, environments structure, entities behavior and both users and agents interactions and activities Chevaillier et al. (2012) and provide tools to directly link a 3D environment and its components to semantic representations Marion et al. (2007). As MASCARET is a meta-model, specific domains are considered as input data. Human domain experts can provide such data. These data remain explicit during the simulation, thus they can serve as a knowledge base for agents playing roles Chevaillier et al. (2012).

In the context of this article, human domain experts are teacher expert and domain data are scenarios. The benefit of MASCARET is that it offers a direct link between scenarios defined by teacher expert and the 3D classroom environment, while keeping semantic (student and teacher roles, behaviors, objects) Marion et al. (2007). Such properties are not present in others solutions Lugin et al. (2019). MASCARET has been largely used and efficient in other 3D applications for coordination between medical staff training Hoareau et al. (2013), firefighters Querrec et al. (2004), aviation management Marion et al. (2007) and procedural learning Le Corre et al. (2014).

Scenarios are implemented through UML activity diagrams containing a sequence of actions. These actions can be either domain actions, like manipulating an object or pedagogical actions, like explaining a resource Buche et al. (2010); Querrec et al. (2004). As vignettes contain sequences of actions with different actors, we were able to integrate these massive scenarios inside a 3D environment using MASCARET Delamarre et al. (2017). It runs with the Unity 3D engine and its scripts are executed at the same frame rate that Unity 3D scripts. Therefore the reaction times of the students to a teacher user input are linked to Unity 3D frame rates. We recorded 25 frame per second with a middle class graphics cards which result in no latency in the reaction time.

Since vignettes were designed using LucidChart, with different shapes used for different meanings, we were able to parse them into activity diagrams using a simple tagging system. For example, diamonds in the vignette, corresponding to a multiple-choice node, are interpreted as decision-merge node in UML. MASCARET uses partitions of activity diagrams to differentiate actions done by different agents. By tagging the acting student in a vignette node, we can attribute an action to be performed by this virtual student (Fig. 6).

The method of translating LucidChart directly into a 3D simulation present 2 main benefits: (1) this method provides a fast way to integrate new classroom situations to IVT-T, i.e. new vignettes respecting the tagging system can be added to the system, provided that students and actions used in the new vignette already exist, without modifying the

current implementation of IVT-T; (2) given the multi-disciplinary nature of IVT-T, this method allows for simple communication between the education expert team and the software engineering team. IVT-T's education experts can share instructional content using their formalism, i.e. the LucidChart diagrams representing the vignette scenarios, and the IVT-T system takes care of interpreting it into a simulation.

4.5. User Interface

The UI of any system plays a major part in how users accept and enjoy interacting with it. As shown in the overview of the system (Fig. 3), users interact with the IVT-T either through the simulator to complete practice sessions or through the website to access pedagogical resources and their simulation logs. In this section, we will present the UI as an interaction pathway between a teacher and IVT-T. First, teachers must log in into the IVT-T website to download the simulator application. No installation is required, teachers can just launch the application which will prompt them to select a classroom, a student and a difficulty level to practice (Fig. 7). At the beginning of a practice, teachers can read a short introduction of a classroom situation before starting the simulation.

The simulator displays the 3D classrooms and the virtual students playing the vignette scenarios. Currently, we identified two main interactions:

- listening and/or reading student virtual students' utterances
- making a decision to direct student behaviors

Video games including narratives and decision making features were surveyed to inspire IVT-T's UI first design which was refined after conducting pilot studies with participants. In the final version, student utterances were presented with a light color font on a dark background along with a portrait picture of the student at the top of the screen. ECTs' utterances used the same layout but were presented at the bottom of the screen.

The decision selection feature (Fig. 8) is also displayed at the bottom of the screen to maintain consistency with ECTs' utterances display. Up to 3 choices can be made at each decision node in the scenario. Choices are displayed as text, associated with a number, describing the action to be performed and showing the dialogue to be communicated to the virtual students. The text is highlighted when hovered by the mouse to indicate the possibility of interaction. Once a choice is selected, it is displayed at the bottom of the screen and teachers need to click on a "Continue" button to see the student reaction to their choices.

Additionally, to enhance the ECTs immersion in the environment,

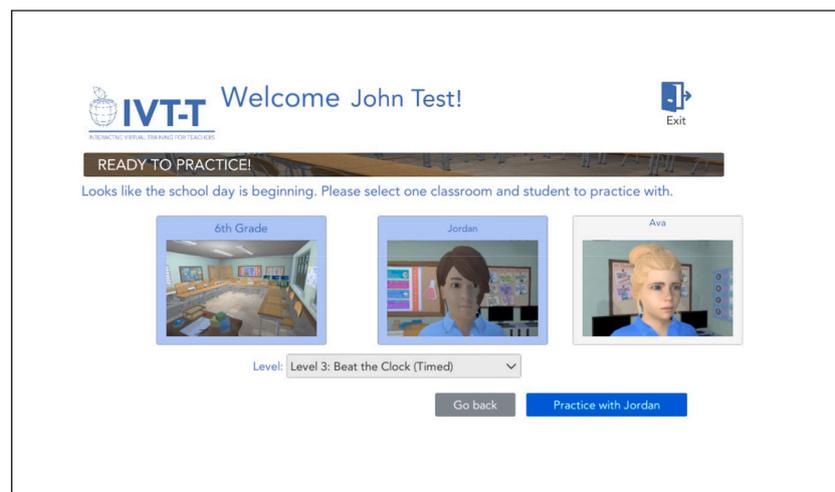


Fig. 7. Selection. Teachers first select the classroom level they want to practice with, then the student and finally the difficulty level.



Fig. 8. Practice. Teachers are presented with three choices to address the student behavior. When practicing at level 3, a timer (top right) indicates to teachers the remaining time to make a decision. If a choice is not selected before the time runs out, the worst choice is automatically selected.

actual children were recorded saying the phrases of the utterances. By using the audio source component in Unity, we emit sounds from a particular location in the classroom. For example, if a student is speaking on the right of the teacher, the user perceives it as so, thus improving spatial realism.

The simulation alternates between students' actions and utterances and teacher's decisions until the end of the scenario is reached. Then teachers are directed to the reflection phase (Fig. 9). ECTs are given the opportunity to reflect on their decisions by selecting a decision to reflect on and which will then prompt a question to guide their reflection.

After the reflection phase ECTs received feedback for each choice they made while interacting with the disruptive virtual student. The feedback is presented on top of a screenshot taken when they clicked on the choices option (Fig. 10). Quantitative feedback (score) is displayed next to each choice, total score is kept on a trophy on the top left. The qualitative feedback is shown in the middle of the screen and provide an assessment of the choice made.

After reading all the feedback, teachers are directed to the storyline summary on the IVT-T website (Fig. 11). ECTs can visualize a storyline summary showing the number of effective decisions made and the strategies used. For each decision, they also have access to the choice they made, the feedback and score they received and the reflection they entered. Storyline summaries for all their simulations are available on the IVT-T website. Additionally, ECTs can review the objectives they need to achieve to gain access to more complex levels.

On the website, ECTs can also access disruptive students' biographies (e.g. *Who are they? What are their relationships with classmates and family?*) and other pedagogical content.

Finally, when they logout of the website, ECTs are asked questions of the Teacher Strategies Questionnaire (TSQ) Webster-Stratton (2005) such as *"How confident are you in managing current behavior problems in your classroom?"*. ECTs can visualize their answers to this questionnaire over time to witness their progress using IVT-T (Fig. 12).

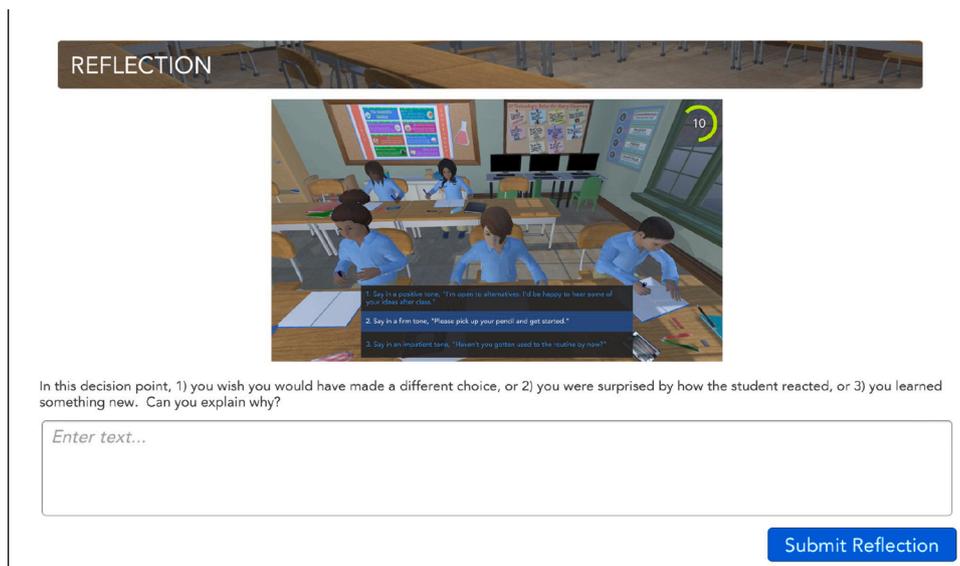


Fig. 9. Reflection. After selecting a decision to reflect on, teachers are invited to answer a question to guide their reflection. A screenshot of the decision is displayed to help teachers consider the choices they had.

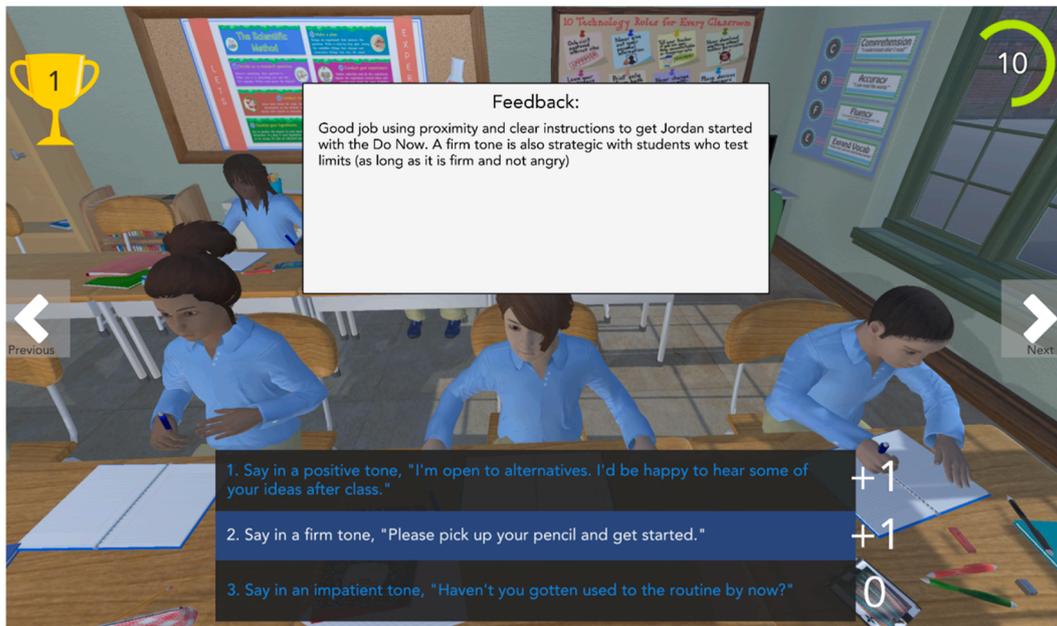


Fig. 10. Feedback. After the reflection phase, ECTs visualize in sequence screenshots of their decisions with corresponding quantitative and qualitative feedback. To move through the sequence of decision screenshots, ECTs use the arrows on the side of the screen.

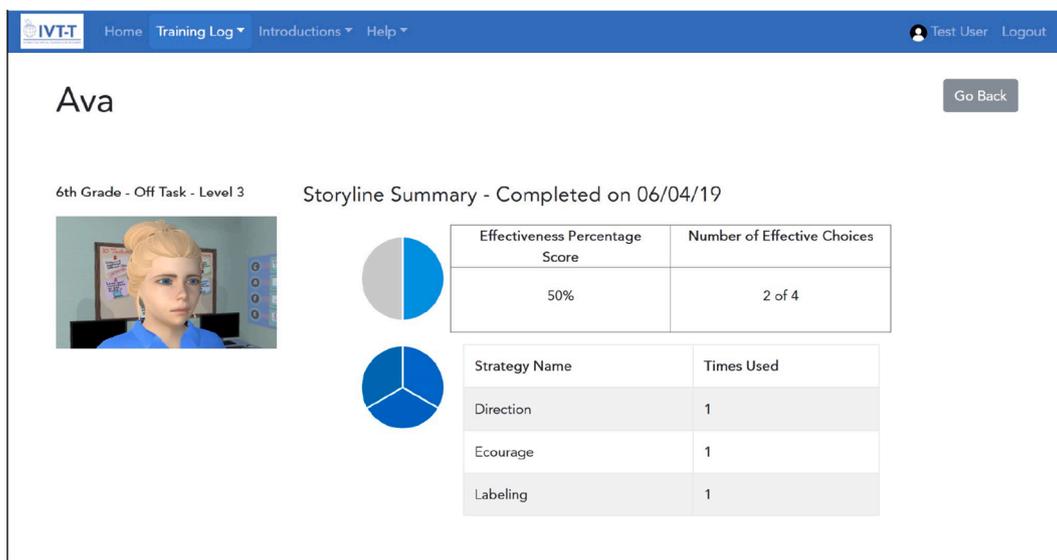


Fig. 11. Storyline Summary. The summary indicates the number of effective choices and the strategies used. Below the summary, users can review decisions and the feedback they received for each simulation they completed.

5. Evaluation

The evaluation of IVT-T was spread over four years, when different requirements were evaluated each year (Table 3). In this section, (1) we first describe the evaluation of vignette scenarios that were rated by experienced teachers to determine their authenticity and realism (Section 5.1); (2) we follow with the evaluation of the realism of the 3D classrooms and virtual student that were rated by the same experienced teachers (Section 5.2); (3) we then present the usability study of the IVT-T system conducted with education majors during year 2 (Section 5.3); (4) we discuss the evaluation of the usage of the IVT-T training sessions which was conducted with practicing teachers (Section 5.4); and finally (5) we present the results of our ongoing study aiming at measuring the effectiveness of the IVT-T approach compared to a traditional approach (control) (Section 5.5).

5.1. Behavioral Fidelity: Vignette scenarios

To deliver realistic and engaging content to ECTs, IVT-T scenarios were assessed in terms of realism, consistence and engagement. A total of 12 vignettes were evaluated by the advisory board: 3 scenario levels for 4 virtual disruptive students. The methods and more comprehensive results can be found in the Shernoff et al. (Shernoff et al. (2018)).

Population. The vignette scenarios were evaluated by the same advisory board who evaluated the classrooms and students and which is composed of retired educators (N = 6) with experience teaching in elementary schools.

Research Protocol. Each vignette was refined based on the observations and feedback collected during the previous evaluation by the advisory board.

Material. A total of 5 vignettes prototypes, for each individual

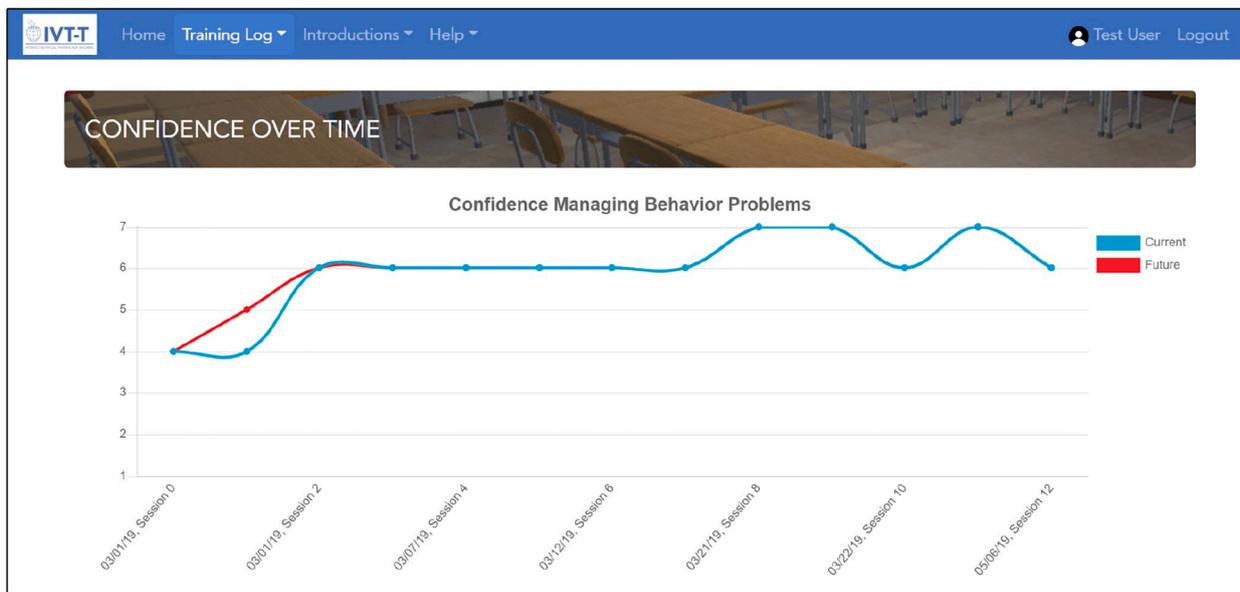


Fig. 12. Confidence. TSQ answers Webster-Stratton (2005). The graph shows the evolution of ECTs’ confidence in managing disruptive behaviors in their classroom currently and in the future.

Table 3
Evaluation of IVT-T for the proposed requirements

Section	Year	Requirement	Type of participant	Evaluation
5.1	1	Behavioral Fidelity	Expert (Experienced teachers)	Vignette realism
5.2	1	Environment Fidelity	Expert (Experienced teachers)	Virtual classrooms and students graphical realism
5.3	2	Usability (Interactivity)	End-users (Education majors)	Usability and UX of IVT-T
5.4	3	Instructional Design: Usage	End-users (Practicing teachers)	Usage of IVT-T training sessions and level up
5.5	4	Instructional Design: Effectiveness (Ongoing)	End-users (Practicing teachers)	Measure of student behaviors in real classroom and observation of teachers use of CBM behavior management (IVT-T vs. Control)

vignette, were presented to the participants in the format of a tree of nodes where the branch splits represent teachers’ decisions and nodes represent actions and dialogue (see left of Fig. 6).

Collected Data. The advisory board provided quantitative and qualitative feedback for each prototype. Logic and realism of behavior and dialogue for the main character, for the non-disruptive students and for the teacher were rated using a 4-point scale (1 = Strongly Disagree, 2 = Disagree, 3 = Agree, and 4 = Strongly Agree) according to the academic level (1st grade and 6th grade). The advisory board also evaluated how engaging the vignette scenario were using the same 4-point scale. Similarly to *VirtualPREX* Dalgarno and Lee (2009) and *SimSchool* Badiee and Kaufman (2015), qualitative feedback was collected and consisted of changes proposition to improve the scenarios.

Vignette scenarios evaluation results. The ratings of the vignettes were positively skewed. Results show that the earlier versions of the vignettes were rated lower than revised versions for each evaluation category (Fig. 13). Vignettes depicting aggressive behaviors (Median = 4.0) were generally rated more engaging than off-task behaviors (Median = 3.5).

Overall realism and logic ratings of the scenario by the experienced teachers are high thus indicating the successful implementation of realistic scenarios.

Participants also provided qualitative feedback, indicating specific sequences of student-teacher interactions that they found particularly strong or in need of improvement. For example, one advisory board member shared: “I think the teacher would have reacted to the phone since most schools do not allow them in the classroom” or “I think the non-disruptive students would react to Jordan’s comment here”. This type of feedback helped directing the refinements in the vignettes and generate more realistic situations.

Evaluation and consecutive refinements of the classroom scenarios resulted in the implementation of realistic classroom behavior thus addressing the behavioral fidelity requirement.

5.2. Environment Fidelity: 3D Classrooms and 3D Virtual Students

To address the environment fidelity requirement, two virtual classrooms (1st grade and 6th grade) and virtual students (15 1st graders and 15 6th graders) were evaluated in terms of authenticity, realism and representativeness. The methods and more comprehensive results can be found in the Shernoff et al. Shernoff et al. (2018).

Population. An advisory board was created to assess the classrooms and students. The board was composed of retired educators (N = 6) with experience teaching in elementary schools.

Research Protocol. A total of 6 classrooms prototypes (for both the 1st grade and 6th grade classrooms) were evaluated by the advisory board. Each prototype was refined based on the observations and feedback collected during the previous evaluation by the advisory board. The evaluation of the virtual students followed the same protocol, however only 4 prototypes were proposed for each of the 30 students.

Material. The advisory board members accessed screenshots of the classrooms and of the students through a secure website.

Collected data. The advisory board provided quantitative and qualitative feedback for each prototype. Regarding the classrooms, feedback focused on physical arrangement (size of the room, desk placement, furniture disposal), wall decorations (bulletin boards, students work, classroom rules), materials and physical appearance (lighting, colors). For the virtual students, advisory board members centered their feedback on specific features of the virtual characters, face, body, hair and clothing. Each characteristic was rated on a 4 point-scale (1 = Poor, 2 =

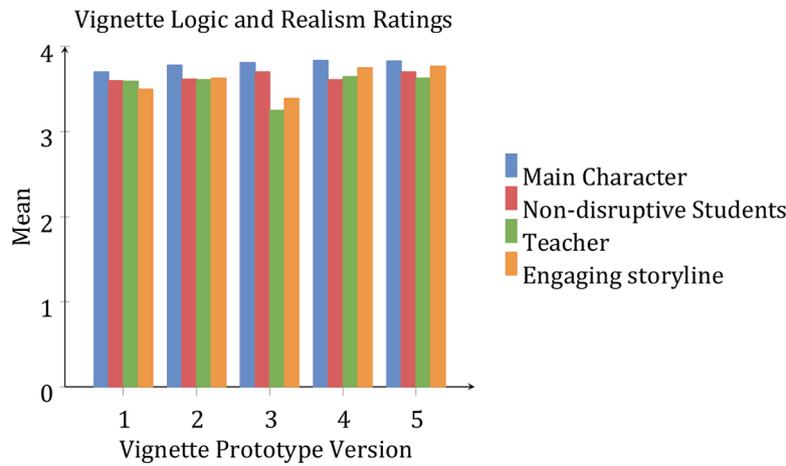


Fig. 13. IVT-T Vignettes logic and realism ratings for the 5 evaluations (1 = Strongly Disagree - 4 = Strongly Agree). Vignettes were refined between each evaluation Shernoff et al. (2018).

Fair, 3 = Good and 4 = Outstanding). Qualitative feedback consisted of recommendations on physical characteristics for each prototype. The metrics used to evaluate IVT-T’s environmental fidelity overlap with qualitative feedback collected on proportionality of classroom furniture and students’ clothes and faces during a study of *VirtualPREX* Dalgarno et al. (2016).

3D Classrooms evaluation results. Quantitative ratings are shown in Fig. 14. Classrooms were evaluated realistic and authentic for the last prototypes compared to earlier prototypes, suggesting that feedback and refinements helped enhance classrooms quality and appearance.

The advisory board also provided qualitative feedback and suggestions regarding each main characteristic. While evaluating the physical arrangements of the first prototype, a participant commented that “desks and chairs still look too nice and shiny - they wouldn’t be in such good condition.”. During the evaluation of the third prototype the same participant specified that “Desk arranged in groups of 4 looks good - desks look slightly more beat up, older and wood looking”, suggesting that the revision brought to the classroom successfully addressed this participant’s comments. When giving feedback on classroom materials, two participants directly suggested to “Add an American flag”. Finally, regarding the physical appearance of the classroom early versions of the prototypes were lacking light realism, “Not much natural light coming in”. This issue was addressed in final classroom versions, “this classroom obviously has a more realistic look, perhaps in part due to better color and light quality”.

Virtual students evaluation results. Fig. 15 shows the mean of the

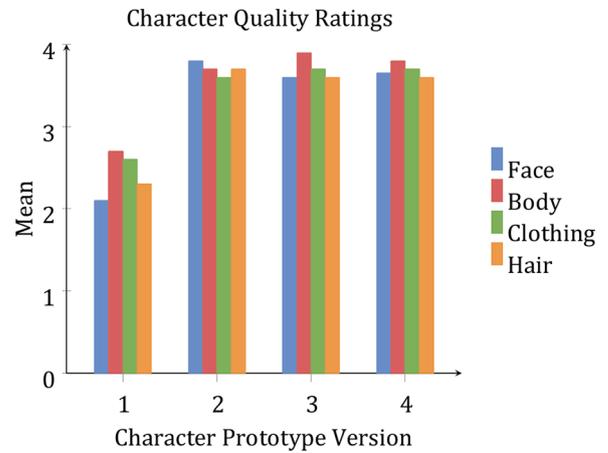


Fig. 15. Quality rating means of the 30 virtual students for the 4 prototypes (1 = Poor - 4 = Outstanding). Virtual students were refined between each evaluation.

quantitative ratings of the advisory boards for each successive virtual students’ prototype evaluations. Refined characters from later evaluations were, overall, rated more realistic that the ones from the first evaluations. The majority of virtual students were evaluated as good to

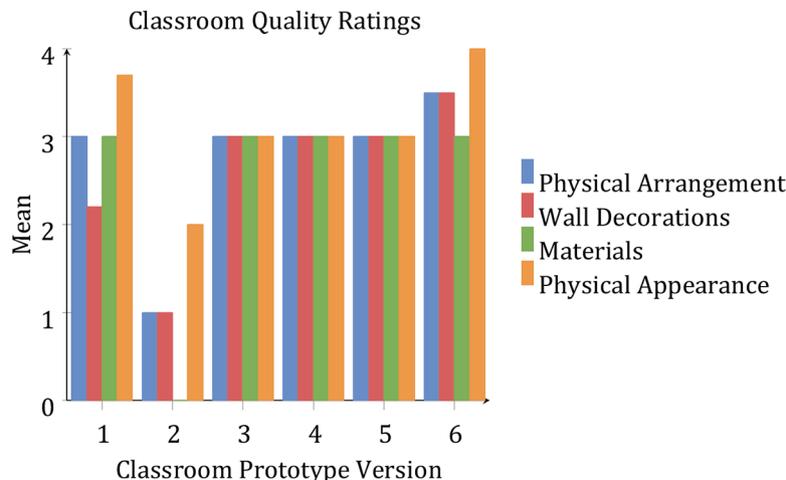


Fig. 14. Quality rating means of the 2 virtual classrooms for the 6 prototypes (1 = Poor - 4 = Outstanding). Virtual classrooms were refined between each evaluation.

outstanding, only 2 characters were rated as fair.

The advisory board provided qualitative feedback to express what they like about the virtual students and to suggest improvements for each avatar feature. For example, when commenting on the face of a first grader, a participant indicated that an avatar looked "Cute and age appropriate" when for another it seems that "Her face looks pinched." and suggested to "Try to soften her out". Regarding the clothing, a participant advised, "Pants need pockets and some other details. The shirt needs buttons down it and maybe a pocket.". When evaluating the hair, suggestions made on the early prototypes such as "Shorten his hair as well" or "Fill out her bangs a little" helped to enhance virtual students' realism as observed by comments made on the last evaluations of virtual students, "I liked that his hair wasn't completely even. It made it seem more realistic."

Qualitative feedback from the advisory board helped improve the quantitative evaluation results of both the classroom and the virtual students. However, as no similar evaluation of graphical realism for VTEs could be found, it is not possible to compare the realism of IVT-T's graphics to other systems. As the impact of graphics quality on the learning outcomes of VTEs are still ongoing research, we encourage existing and future VTEs to proceed to similar graphics evaluation in order to provide comparative values to better study the effect of graphics quality for VTEs.

5.3. Interactivity: Usability and UX

We designed a study to assess IVT-T usability and User Experience (UX). Usability refers to the ease of use and the learnability of the system while UX refers to a person's overall perception of the system. As usability considers the pragmatic aspect, i.e. *how long does it take to achieve a task?*, UX is more related to users' feeling, i.e. *does users like to use the system?*. A secondary objective was to identify issues detrimental to the global utilization of the system. The evaluation focuses on IVT-T's learnability and efficiency. The methods and more comprehensive results can be found in the Shernoff et al. [Shernoff et al. \(2020b\)](#).

Population. Education majors (n=7, 7 female) were recruited from a school of education to participate to this formative evaluation. Criteria for recruitment included the interest of working in elementary schools and an academic level of senior or graduate student for them to have enough experience and background to provide compelling feedback.

Research protocol. All participant interacted with the system individually. After completing the informed consents form, participants were provided a list of tasks to complete using the system. A standardized CTA [Cooke \(2010\)](#) was used during the overall interaction, i.e. participants were instructed to verbalize their thoughts while interacting with the system. After completing all the tasks, a semi-structured interview took place interview to assess satisfaction, ease of use and to gather suggestions to improve the IVT-T system. Finally, the participants completed three questionnaires: (1) Gaming/Computer Experiences Survey Adapted from [Ijsselstein et al. \(2013\)](#), (2) Questionnaire for User Interface Satisfaction (QUIS) [Chin et al. \(1988\)](#) and (3) System Usability Scale (SUS) [Brooke \(1986\)](#). Participants were also asked to provide basic demographic information. Each session lasted approximately two hours (one hour interacting with the system and one hour completing interview research measures).

Material. IVT-T was presented to participants on a desktop computer equipped with an eye tracking device, a webcam and a microphone. Interactions with the system were done using the mouse and the keyboard.

Collected Data. For each participant, two videos were recorded during the interaction with the system; (1) the eye tracking video (which also recorded audio, including the CTA) (2) the webcam video. Time to complete each task and number of errors encountered were manually recorded by observers in the room. Finally, the semi structured interviews were audio-taped, and observers could take notes for each user in order to identify key issues or suggestions to improve the system. Data from questionnaires were also recorded. The Gaming/Computer

Experiences survey explore whether teachers gaming experience influences use of technology by asking questions about gaming habits and gaming experiences. During a usability study of 3B [Lugrin et al. \(2016\)](#), usability was measured through a custom questionnaire on Technology Acceptance, using questions such as, "I could easily explain the system to my colleague." For the evaluation of IVT-T we resorted to well accepted self-report questionnaires, QUIS and SUS, which provide standardized measures to collect quantitative data. Participant rated the 27 items of the QUIS on 10-point scale (e.g., 0 = hard to 9 = easy, 0 = confusing to 9 = very clear, 0 = rigid to 9 =flexible), thus evaluating quality and satisfaction with human-computer interfaces. The SUS measuring system usability uses a 5-point Likert scale ranging from 1 (strongly disagree) to 5 (strongly agree) for 10 items.

5.3.1. Quantitative Results

The game and computer experience survey indicated that two participants never played a video game before using IVT-T. Out of the seven, only two participants consider themselves as gamers, the other five play less than an hour a week. Based on the result of this survey, we identified that most of the participant were not gamers and thus would be able to provide interesting feedback on the usability of the IVT-T system which is meant to use with population not necessarily technology savvy.

The QUIS results were normally distributed (Shapiro-Wilk test $p = 0.159 > 0.05$) and are shown in [Fig. 16](#). The QUIS indicate that, overall, users reacted well to the system (M=7.20, SD = 0.83), revealing that the IVT-T system provide an easy to use experience. Organization and presentation of the information were also well perceived by the participants (Screen (M=7.18,SD=1.56)). Participants were satisfied with the system capabilities (M=6.87,SD=2.05) but results indicated issues with the speed of the system as well as the possibility to correct mistakes. However, regarding the use of terminology and system information, such as use of terms throughout system, terminology related tasks or display error and progress were rated very satisfying by the participants (M=7.73,SD=0.91). Finally, the learning of the system was rated as outstanding (M=8.83,SD=0.59) indicating that IVT-T system is straightforward to use and does not necessitate training in order to use it.

The results of the SUS questionnaire were also normally distributed (Shapiro-Wilk test $p = 0.947 > 0.05$) and confirmed the results of the QUIS, as user thought the system was easy to use (M=4.43, SD=0.54) and that they felt very confident using the system (M=4.57, SD=0.54). Moreover, they did not feel like they needed to learn a lot of things before they could get going with the system (M=1.57,SD=0.78) neither they think that they would need the support of a technical person to be able to use the system (M=1.29,SD=0.49).

5.3.2. Qualitative Results

The semi structured interview corroborated questionnaire results. Participants generally found the system easy to use and straightforward, "I think the program itself was very easy to understand. Understanding how to maneuver and what to do was easiest for me.". Even for participant with very little experience in gaming thought IVT-T was usable: "It was very easy to navigate, so I liked that a lot because I'm not very tech savvy. I found it very easy to use.". As participants were able to interact with IVT-T without facing major usability issues, participants feedback was mostly concentrated on the simulator's content (classrooms and students appearances and behavior realism). Therefore a qualitative content analysis of the semi structured interviews was conducted for the realism of the graphics, audio, and behaviors [Hsieh and Shannon \(2005\)](#). Users comments were classified into three categories: positive impressions, negative impressions and recommended changes.

Regarding graphics and audio realism, two main themes appeared for positive impressions. IVT-T accurately represented virtual students and classrooms according to their academic level, and the voice recordings were able to enhance the realism of the situations. One

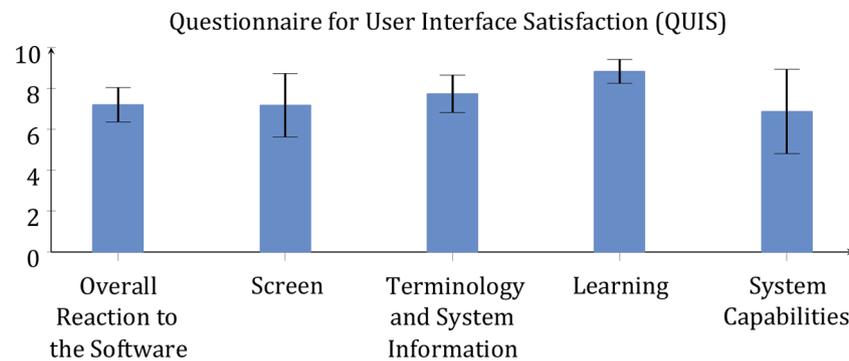


Fig. 16. QUIS is comprised of six subscales. Each of these scales includes between four and six items, rated on a 10-point scale (0 = hard to 9 = easy, 0 = confusing to 9 = very clear, 0 = rigid to 9 = flexible). Items in each subscale are averaged to compute a mean score for each subscale.

participant commented: (No, it was very realistic in my opinion. That's very pre-teen of him like to be you know like trying to be difficult almost like the things he was saying were like just being difficult, trying to distract the teacher and to get attention. At the same time all these things were very encompassing of a sixth grade personality."). A theme that emerged from negative impressions for graphics and audio realism was the lack of details in the classroom. More details were added to the classrooms after this evaluation (e.g. clutters on the students' desk). Finally, participants recommended to integrate the ability for the virtual students to display facial expressions.

For the behavioral realism of the students, positive impressions focused on the realism of the classroom situations provided by IVT-T. One participant mentioned that the display of attitudes by the disruptive student made the situations realistic: "He [Jordan] seemed really realistic. You know, wanting to take up space in the classroom, it's like, well, I'm already late. How much worse can it be? It felt familiar". And I think that makes it more realistic". A theme that emerged from the negative impressions regarding the realism of the vignettes was that only one student was disruptive. "what would happen when Jordan is disruptive and then another student becomes disruptive too?" The vignettes were adapted to better reflect the spread of a disruptive behaviors to neighbors students. Participants also suggested features to integrate to future versions of IVT-T, such as to providing the opportunity to see the simulation of what would happen if a trainee were to make choices that always escalate the situation, or choices that always de-escalate the situation.

5.4. Instructional Design: Usage

The objective of this study was to evaluate the instructional design and the usage of IVT-T with practicing K-8 teachers (Evaluation of IVT-T 3.0 in Fig. 2). The methods and more comprehensive results can be found in Shernoff et al. Shernoff et al. (2020a).

Population. A sample of practicing K-8 teachers ($n=27$, 22 females) were recruited from three moderately-sized urban schools.

Research protocol. Participants were given access to IVT-T for an eight week period in complete autonomy. Participants were asked to use IVT-T for 45 to 60 minutes per week.

Material. Participants used IVT-T on their personal computers.

Collected Data. The system recorded logs of every IVT-T training session (time spent practicing with IVT-T, training session phases completed, number of level up). An IVT-T training session goes through the following sequence: (1) Practice; (2) Replay; (3) Reflection; and (4) Feedback. However, users can decide to exit the session at any time, a training session is recorded if at least the Practice phase is completed, i.e. the teacher reached an end node in the vignette. Teachers also completed rating scales before and after using IVT-T to assess perceived efficacy with technology, along with ratings of usefulness, ease of use, and plans to use simulation training in the future Kao et al. (2014) with response options ranging from 1 - Strongly Agree to 5 - Strongly Disagree.

Additionally, similarly to *SimInClass* Kelleci and Aksoy (2020), 18 teachers participated in digitally recorded focus groups. Two focus group sessions of 60 minutes were conducted with 9 teachers in each. Teachers earned school supplies and professional development credits for participating in the focus groups.

5.4.1. Quantitative results

Over the eight weeks period, the majority of teachers in the sample (81%) logged in and used the system at least once during the 8 weeks. Nineteen percent of users were considered heavy users (used the system, on average, more than 45 minutes), 30% were medium users (used IVT-T between 22.5 and 44 minutes per week), 33% were mild users (less than 22 minutes per week on average). There were 19% of users who had access to the system but did not use it at all. The weekly usage of IVT-T is shown in Fig. 17.

Of the 208 practice sessions started by the teachers, 68% were completed, i.e. teachers went through all the training phases (practice, replay, reflect, and feedback). This result indicates that teachers had an interest in completing the different phases of the training sessions. Additionally, 7 teachers (26%) completed the objectives to access level 3 vignettes and 34 practice sessions at level 3 were completed. Ratings of overall perceived efficacy with technology, ease of use, and plans to use simulation training in the future among teachers using IVT-T 3.0 over the 8 weeks increased from a mean of 32.2 ($SD = 8.9$) to 36.2 ($SD = 4.6$). This was considered a statistically significant increase ($p = .000$).

5.4.2. Qualitative results

Two focus groups ($n = 9$ for each group) were conducted after the eight weeks period and a qualitative content analysis was conducted on their transcripts to extract major themes on the use of IVT-T. We were able to identify four main themes: (1) finding the time to practice with IVT-T, (2) adjusting the amount of human guidance to teachers, (3)

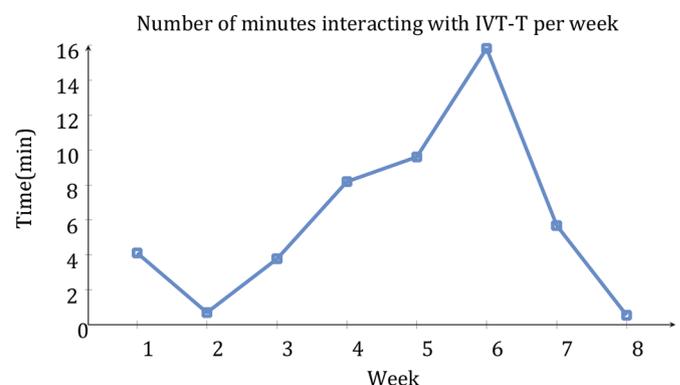


Fig. 17. Total number of minutes interacting with IVT-T during the eight weeks period in Year 3.

instructional design supported and maintained teacher engagement, and (4) IVT-T is more interactive than traditional training.

Teachers had a hard time finding time in their schedule to practice with IVT-T *"its just more finding the time with a life thats busy and squeezing it in"*. Teachers were not able to dedicate a regulate time to log in into IVT-T *(I usually did it Sunday nights after lesson plans, any time at night usually. Before I went to bed)*.

Some teachers liked the human guidance during their first sessions with IVT-T: *I think thats when we met as a group and you showed everybody how to use IVT and everybody participated and everybody practiced scenarios, then when we do it ourselves its easier*. At the beginning of the study, experimenters organized a one-time sessions to assist teachers in login in the IVT-T website, and complete their first IVT-T training session. During this session some teachers *felt really rushed* and would have like more time to practice while being assisted. Organizing meetings with teachers needing more assistance during the first two weeks of the study could have benefited an increase the in IVT-Ts usage with these teachers. However, as identified during these focus groups and by other studies [Dawson and Lignugaris/Kraft \(2017\)](#), teachers schedules are hard to manage.

Nevertheless, other teachers said that more meetings to get accustomed to IVT-T are not *"necessary"* and they autonomously learned to use IVT-T *("You learn it through exploring."* *"You really don't start to get the feel of IVT until you really use it."*). The differences in teachers approach to a new tool to learn CBM underline the complexity faced by VTE-T as there is no one size fits all.

Teachers also noted some glitches with the system *"I was trying to figure out so many glitches I wanted to finish it but I just couldnt"*. As IVT-T was still in development during the year this study took place, and given the massive size of IVT-T's vignettes (approx. 600 nodes per vignette), the prototype IVT-T 3.0 used was not bug-free. However, the IVT-T logging system allowed to pinpoint where in the vignette glitches were encountered and helped the development team to address these issues for the final version of IVT-T. Additionally comments like *"One is like my computer K glitching and the other part is the system glitching"* underline the importance of technological aspect of the Autonomy requirement. Not all teachers have access to a computer with enough graphical power to run a high definition 3D simulations. To reach a high number of teachers with a VTE-T and provide an enjoyable and playable, the system requirement must be low enough to allow most computers to run the simulation.

During the focus group, teachers reported that the interaction design of IVT-T for the decision making process (the display of the three choices and the continue button to observe the student reaction to the selected choice) help maintain their engagement during practice: *"I do like that you can hit the continue button and you are prompted to hit the continue button because obviously it kind of holds us accountable to keep watching the program.. because you know K You could just zone out. But in the moment, Im like, oh wait, I did not, I have to hit the continue button."* Some teachers liked to have to read the teachers options *("I liked reading it")* as it helped them mentally representing the classroom situation: *"I actually don't mind reading something because it really brings the scene to life"*.

Finally, a main theme that emerged from the focus groups was that some teachers preferred an approach like IVT-T compared to traditional training approaches: *"Why I prefer this [IVT-T] more, is that it attended to my own dominant learning style, along with added, Im very visual, so, after a while, in a workshop, I glass over."* Some teachers found the interactive nature of IVT-T more engaging *("This is a lot more fun"*, *"I like the hands-on element."*) and contrasted IVT-T with passive training workshops: *"You know so often we go to workshop, and its like, look at this video, look at this power point, look at this, read, read that. And Im like I dont want to read that, I dont care about that, Im doing it my way anyway so I dont care. But this gave me something to do."*

Overall, the results of this study indicate that IVT-T is a feasible way to deliver CBM training to some teachers, and that it offers an effective engaging tool to provide teachers with a hands-on experience.

5.5. Instructional Design: Effectiveness (Ongoing)

The objective of this study is to evaluate the instructional design effectiveness of IVT-T with practicing K-8 teachers (Evaluation of IVT-T 4.0 in [Fig. 2](#)). However, so far, only the data on teachers usage of IVT-T is available.

Population. A sample of practicing K-8 teachers (n=26) were recruited from six elementary schools.

Research protocol. Participants were given access to IVT-T for a 20 weeks period in complete autonomy. During a briefing section at the start of the period participants were given accounts to access IVT-T and were guided to complete their first simulation and visualize their logs on the website.

Material. Participants used IVT-T on their personal computers.

Collected Data. The system recorded logs of every IVT-T training session (Scores, reflection entered, and feedback received). In this study, the Replay phase of the IVT-T training session was optional and to level up, teachers had to complete two unique storylines (instead of four in year 3) with an effectiveness score of 65% or more (instead of 80% or more in year 3). An IVT-T training session is considered complete if the feedback for all the decision have been received by the user.

5.5.1. Quantitative Results

During the 20 weeks period, a total of 1,064 IVT-T training sessions were recorded. Of the recorded IVT-T training sessions at level 2 and level 3, 87.8% were fully completed ([Fig. 18](#)). This result is an improvement from the study conducted in year 3 ([Section 5.4](#)) which only had 68% of completed training sessions. We observe that teachers were able to level up and access the level 3 vignettes, 48.5% of practice sessions played were at level 3 (only 16.3% of level 3 played during year 3 ([Section 5.4](#))). Teachers chose to always skip the optional replay. During year 3, the replay phase was mandatory and teachers spent on average 3 minutes and 22 seconds in the replay phase.

The number of teachers, the length of the time period to practice with IVT-T and the training sessions of this study were different from the study presented in [Section 5.4](#), therefore we differentiated usage rates between teachers with the number of storyline played per week. Of the 26 participants, 19% were heavy users (teachers playing more than 5 storylines per week on average), 19% were medium users (teachers playing at least once storyline per week on average): 62% mild users (teachers playing less than one storyline per week on average). All teachers used IVT-T at least once. The number of practice sessions played each week is shown in [Fig. 19](#)

5.5.2. Qualitative Results

A qualitative content analysis was conducted on the reflections

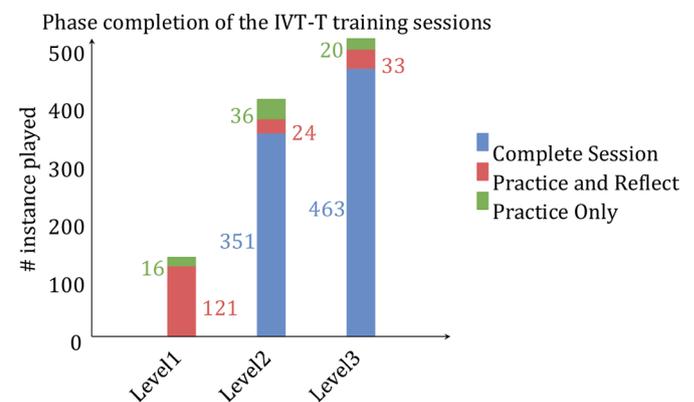


Fig. 18. Completion of the different phases of IVT-T training sessions. A complete session is counted when the teacher trainee finishes all the phases (Practice, Replay (Optional), Reflect, Feedback). Level 1 vignettes do not include the feedback phase.

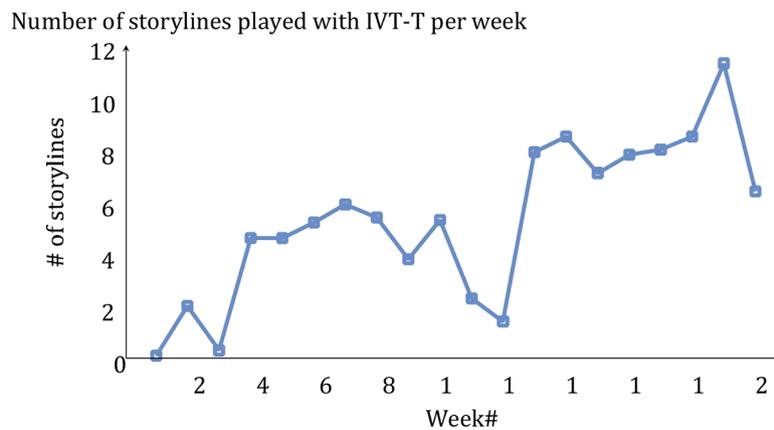


Fig. 19. Total number of storylines played with IVT-T during the 20 weeks period in Year 4.

entered by teachers to extract major themes on teachers use of IVT-T. Only 142 reflections (13%) were left blank. A theme that emerged from the reflection was the connection between IVT-T and real classroom experiences, as seen from sample quotes from users: “Great decision. Made the student happy and it did not disrupt the class. I have done this in my class, in the past, and still do it currently when the situation arises.”, “This tactic works very well when I use it in my classroom. It is never good to go back and forth with a child who is already exhibiting unruly behavior.”). This type of reflection confirms the realism of IVT-T scenarios and simulation and shows that teachers create parallels between the simulation and an actual classroom situation.

Moreover, some reflections seem to indicate that teachers were learning from their practice with IVT-T. For example, a participant wrote: “As a new teacher I come to find it is not conducive to reiterate negative behavior to students, as if they do not already know what they did wrong. As with Jordan, there was not a need to remind him that he was already late, because that exasperated the problem and made him more angry. The best solution is to watch intently and allow time for Jordan to get himself together and join the rest of the class.”. Another teacher mentioned the following: “I have learned not to focus one the student but on the behavior. Avoid to call the student name when they are not cooperating.”.

Teachers also mentioned the benefit of using empathy in the classroom: “By showing empathy to Jordan, the teacher opens a door for Jordan to feel more confident with the teacher.” and “It helps to show empathy to students without getting too personal in front of the entire class.”.

Finally, we observed that teachers also used IVT-T to explore the effect of different behavior management strategies: “I picked these choices to see what situations can escalate to if they are not handled properly.”.

These results show that the IVT-T training sequence was efficiently used by teachers and they support the necessity to integrate teachers reflection to the instructional design, as it increases the transfer of learning Merrill (2009) and provide good insight on the use of the system.

6. Discussion

The development of IVT-T was driven by the requirements established from the previous work in VTEs for ECTs’ training. Each of them includes features that we assessed as necessary to build an efficient training system for teacher and addresses limitations of previous systems:

1 - Behavioral Fidelity: An advisory board composed of experienced teachers evaluated the vignette scenarios content which was refined iteratively until assessed as realistic. IVT-T is the only classroom simulator which validated the content of its scenarios. The realism of IVT-T’s content was also pinpointed during the usability evaluation.

2 - Environment Fidelity: IVT-T includes a 3D environment constituted

of two classrooms validated as realistic by educators with experience in elementary schools in terms of physical arrangement, materials decorations and physical appearance. Thirty virtual students impersonating fifteen 1st graders and fifteen 6th graders presenting a realistic number of students per class, compared to the TeachLive simulator which presented a classroom of 6 students. The advisory board evaluated the virtual students as authentic and representative in terms of face, body, clothing, and hair.

3 - Instructional Design: IVT-T provides ECTs with learning experiences where they are encouraged to practice and reflect. Moreover, by enabling ECTs to replay simulations of their IVT-T sessions, ECTs can learn from their mistakes and find way to improve their skills. Additionally, IVT-T offers different levels of difficulties for ECTs to practice their CBM skills. Objectives must be attained in order to access more complex levels, offering an evolution of difficulty that ECTs can follow at their own pace. Finally, IVT-T invites ECTs to reflect on their choices, so they can assess their own performance. As shown by Reigeluth et al. Merrill (2009), integrating reflection to the instruction increase the transfer of learning. Finally, ECTs receive automatically generated explicit feedback on the choices they have made which result in better strategy retention and transfer of learning Richey et al. (2011); Tracey et al. (2014). A preliminary study of IVT-T showed that teachers efficiently use the IVT-T training sequence.

4 - Autonomy: IVT-T is completely autonomous (no human is needed to run the system) from practice to feedback, unlike other systems necessitating human operators. IVT-T website guarantees an online access and low-technology requirements. IVT-T thus facilitates ECTs getting practice and receiving feedback about their CBM skills.

5 - Interactivity: Interactions with the system have been studied to yield a self-explained and effective UI. The usability and UX study showed IVT-T’s efficiency and learnability.

6 - Scalability: Using the MASCARET framework, IVT-T can quickly and easily integrate new classroom vignette scenarios and thus can present a variability of classroom situations for ECTs to practice and reflect on.

These requirements, in addition to guide the development of the IVT-T classroom simulator, also raised open-ended research questions.

The impact of graphics realism on the transfer of learning is still an ongoing research. Some posit that high-quality graphics will generate a greater sense of immersion and engagement Bossard et al. (2008); Dalgarno and Lee (2009) which will result in more transfer of learning. Whereas others argue that, too much fidelity would fall in the uncanny valley and would thus be detrimental for learning Wages et al. (2004).

Determining the degree of realism required to optimize the learning, as well as improving the efficiency of VTEs, would help determine how many resources should be allocated to graphics, as high-fidelity graphics are costly and time consuming.

The autonomy of IVT-T allows ECTs to practice and receive feedback

on their own. The IVT-T students' behaviors are scripted by the vignette scenarios, however resorting to socially intelligent agents could provide more adaptability to the users, i.e. the difficulty of the situation could evolve at runtime depending on the ECTs' performance. Furthermore, it would introduce variability from one simulation to the other and therefore keep users engaged by removing simulation repetitiveness. The survey of existing work in the domain revealed that only one classroom simulator resorted to a model-based approach. However, the difference of impact on learning outcomes between these two approaches or a combination of both has not been studied.

TeachLive and 3B simulators rely on human instructors to provide feedback to teachers, however, it greatly reduces teacher accessibility and autonomy. IVT-T is built on an ITS architecture, integrating an intelligent tutor in IVT-T could provide more personalized feedback, as a human instructor would, while maintaining IVT-T's autonomy.

7. Conclusion

This article presented IVT-T, a 3D classrooms simulator to provide active training for ECTs to enhance their behaviors management skills. ECTs can interact with realistic virtual students exhibiting disruptive behaviors at two different academic level 1st grade and 6th grade.

We established requirements for the development of VTE-Ts in order to design and implement an efficient simulator and address issues revealed by existing systems such as environment fidelity, behavioral fidelity, and autonomy.

During the implementation, efforts were concentrated on design and prototyping of graphics (virtual students and classrooms) and on vignette scenarios which were iteratively evaluated and refined by education expert and experienced teachers. User Interfaces have also been designed, evaluated and refined, in order to provide an easy to learn and easy to use experience and avoid frustration that could arise from ECTs struggling to use the system (which as pointed out earlier, has been observed with other systems)

The final challenge is to determine if IVT-T leads to better or more rapid acquisition of behavior management skills than the costly trial-and-error approach ECTs are taking in their classroom. A quasi-experimental design will be used to compare rates in change of slope from pre- to post-test between ECTs randomly assigned to traditional learning and to traditional learning with IVT-T.

The final challenge is to determine if IVT-T leads to better or more rapid acquisition of behavior management skills than the costly trial-and-error approach ECTs are taking in their classroom. A quasi-experimental design will be used to compare changes in teacher effectiveness and student disruptive behavior when comparing teachers with and without access to IVT-T.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R305A150166 to the Rutgers, the State University of New Jersey. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education. Alban Delamarre is supported by a Florida International University (FIU) Dissertation Year Fellowship.

References

- , 2019. The ux book (second edition). In: Hartson, R., Pyla, P. (Eds.), 2019. The UX Book (Second Edition), Second edition. Morgan Kaufmann, Boston, pp. xxv–xxvi.
- Attebery, A., Loeb, S., Wyckoff, J., 2017. Teacher churning: Reassignment rates and implications for student achievement. *Educational Evaluation and Policy Analysis* 39 (1), 3–30.
- Badiee, F., Kaufman, D., 2015. Design evaluation of a simulation for teacher education. *Sage Open* 5 (2).
- Bailenson, J.N., Swin, K., Hoyt, C., Persky, S., Dimov, A., Blascovich, J., 2005. The independent and interactive effects of embodied-agent appearance and behavior on self-report, cognitive, and behavioral markers of copresence in immersive virtual environments. *Presence: Teleoperators & Virtual Environments* 14 (4), 379–393.
- Bangor, A., Kortum, P.T., Miller, J.T., 2008. An empirical evaluation of the system usability scale. *Intl. Journal of Human-Computer Interaction* 24 (6), 574–594.
- Barmaki, R., Hughes, C.E., 2015. Providing real-time feedback for student teachers in a virtual rehearsal environment. *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction*, pp. 531–537.
- Barot, C., Lourdeaux, D., Burkhardt, J.-M., Amokrane, K., Lenne, D., 2013. V3s: A virtual environment for risk-management training based on human-activity models. *Presence* 22 (1), 1–19.
- Bates, J., et al., 1994. The role of emotion in believable agents. *Communications of the ACM* 37 (7), 122–125.
- Bossard, C., Kermaec, G., Buche, C., Tisseau, J., 2008. Transfer of training in virtual environments: A new challenge. *Virtual Reality* (12), 151–161.
- Bowman, D.A., Gabbard, J.L., Hix, D., 2002. A survey of usability evaluation in virtual environments: classification and comparison of methods. *Presence: Teleoperators & Virtual Environments* 11 (4), 404–424.
- Brenton, H., Gillies, M., Ballin, D., Chatting, D., 2005. The uncanny valley: does it exist. *Proceedings of conference of human computer interaction, workshop on human animated character interaction*. Citeseer.
- Brooke, J., 1986. System usability scale (sus): a quick-and-dirty method of system evaluation user information. Reading, UK: Digital Equipment Co Ltd 43.
- Buche, C., Bossard, C., Querrec, R., Chevallier, P., 2010. PEGASE: A generic and adaptable intelligent system for virtual reality learning environments. *International Journal of Virtual Reality* 9 (2), 73–85.
- Chellali, A., Mentis, H., Miller, A., Ahn, W., Arikatla, V.S., Sankaranarayanan, G., De, S., Schwaitzberg, S.D., Cao, C.G., 2016. Achieving interface and environment fidelity in the virtual basic laparoscopic surgical trainer. *International journal of human-computer studies* 96, 22–37.
- Chevallier, P., Trinh, T.-H., Barange, M., De Loo, P., Devillers, F., Soler, J., Querrec, R., 2012. Semantic modeling of virtual environments using mascaret. 2012 5th Workshop on Software Engineering and Architectures for Realtime Interactive Systems (SEARIS). IEEE, pp. 1–8.
- Chin, J.P., Diehl, V.A., Norman, K.L., 1988. Development of an instrument measuring user satisfaction of the human-computer interface. *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM, pp. 213–218.
- Christensen, R., Knezek, G., Wood, T.T., Gibson, D., Tyler-Wood, T., Gibson, D., Knezek, G., Gibson, D., 2011. SimSchool: An online dynamic simulator for enhancing teacher preparation. *International Journal of Learning Technology*.
- Collum, D., Christensen, R., Delicath, T., Knezek, G., 2020. Measuring changes in educator bias in a simulated learning environment. *Society for Information Technology & Teacher Education International Conference*. Association for the Advancement of Computing in Education (AACE), pp. 496–502.
- Cooke, L., 2010. Assessing concurrent think-aloud protocol as a usability test method: A technical communication approach. *IEEE Transactions on Professional Communication* 53 (3), 202–215.
- Creemers, B.P.M., 1994. Effective instruction as a basis for effective education in schools. *Tijdschrift Voor Onderwijsresearch* 19 (1), 3–16.
- Creswell, J.W., Plano Clark, V., 2011. Choosing a mixed methods design. *Designing and conducting mixed methods research* 2, 53–106.
- Crone, L.J., Teddlie, C., 1995. Further examination of teacher behavior in differentially effective schools: Selection and socialization process. *Journal of Classroom Interaction* 30 (1), 1–9.
- Dalgarno, B., Gregory, S., Knox, V., Reiners, T., et al., 2016. Practising teaching using virtual classroom role plays. *Australian Journal of Teacher Education* (Online) 41 (1), 126.
- Dalgarno, B., Lee, M.J., 2009. What are the learning affordances of 3-D virtual environments? *British Journal of Educational Technology* 41 (1), 10–32.
- Dawson, M.R., Lignugaris/Kraft, B., 2017. Meaningful practice: Generalizing foundation teaching skills from the teachlive to the classroom. *Teacher Education and Special Education* 40 (1), 26–50.
- De Freitas, S., Oliver, M., 2006. How can exploratory learning with games and simulations within the curriculum be most effectively evaluated? *Computers & education* 46 (3), 249–264.
- De Freitas, S., Rebollo-Mendez, G., Liarakis, F., Magoulas, G., Pouloussis, A., 2010. Learning as immersive experiences: Using the four-dimensional framework for designing and evaluating immersive learning experiences in a virtual world. *British Journal of Educational Technology* 41 (1), 69–85.
- Debattista, K., Bashford-Rogers, T., Harvey, C., Waterfield, B., Chalmers, A., 2017. Subjective evaluation of high-fidelity virtual environments for driving simulations. *IEEE Transactions on Human-Machine Systems* 48 (1), 30–40.
- Dede, C., 2009. Immersive interfaces for engagement and learning. *science* 323 (5910), 66–69.
- Delamarre, A., Buche, C., Polceanu, M., Lunn, S., Ruiz, G., Bolivar, S., Shernoff, E., Lisetti, C., 2017. An Interactive Virtual Training (IVT) simulation for early career

- teachers to practice in 3D classrooms with student avatars. *FLAIRS 2017 - Proceedings of the 30th International Florida Artificial Intelligence Research Society Conference*.
- Denton, C.A., Hasbrouck, J., 2009. A description of instructional coaching and its relationship to consultation. *Journal of Educational and Psychological Consultation* 19, 150–175.
- Dieker, L.A., Hynes, M.C., Hughes, C.E., Hardin, S., Becht, K., 2015. TLE TeachLivE: Using Technology to Provide Quality Professional Development in Rural Schools. *Rural Special Education Quarterly*.
- Drettakis, G., Roussou, M., Reche, A., Tsingos, N., 2007. Design and evaluation of a real-world virtual environment for architecture and urban planning. *Presence: Teleoperators and Virtual Environments* 16 (3), 318–332.
- Evertson, C.M., Weinstein, C.S., 2013. *Handbook of classroom management: Research, practice, and contemporary issues*. Routledge.
- Gao, Z., Yu, Y., Zhou, Y., Du, S., 2015. Leveraging two Kinect sensors for accurate full-body motion capture. *Sensors* 15 (9), 24297–24317.
- Gerbaud, S., Mollet, N., Ganier, F., Arnaldi, B., Tisseau, J., 2008. GVT: A platform to create virtual environments for procedural training. *Proceedings - IEEE Virtual Reality*, pp. 225–232.
- Geuss, M., Stefanucci, J., Creem-Regehr, S., Thompson, W.B., 2010. Can I pass? using affordances to measure perceived size in virtual environments. *Proceedings of the 7th Symposium on Applied Perception in Graphics and Visualization*, pp. 61–64.
- Gibson, D., 2011. Modeling emotions in simulated learning. *Standards in Emotion Modeling*.
- Gregory, S., Dalgarno, B., Crisp, G., Reiners, T., Masters, Y., Dreher, H., Knox, V., 2013. VirtualPREX: Innovative assessment using a 3D virtual world with pre-service teachers. *Office for Learning and Teaching*.
- Grossman, P., McDonald, M., 2008. Back to the future: Directions for research in teaching and teacher education. *American Educational Research Journal* 45, 184–205.
- Grossman, P., McDonald, M., 2008. Back to the future: Directions for research in teaching and teacher education.
- Gupta, S., Anand, D., Brough, J., Schwartz, M., Kavetsky, R., 2008. Training in virtual environments: A safe, cost effective, and engaging approach to training.
- Hayes, A., Johnson, K., 2019. Cultural embodiment in virtual reality education and training: A reflection on representation of diversity. *Foundations and Trends in Smart Learning*. Springer, pp. 93–96.
- Hayes, A.T., Hardin, S.E., Hughes, C.E., 2013. Perceived presences role on learning outcomes in a mixed reality classroom of simulated students. *International Conference on Virtual, Augmented and Mixed Reality*. Springer, pp. 142–151.
- Hayes, A.T., Straub, C.L., Dieker, L.A., Hughes, C.E., Hynes, M.C., 2013. Ludic learning: Exploration of the teachlive and effective teacher training. *International Journal of Gaming and Computer-Mediated Simulations (IJGCMS)* 5 (2), 20–33.
- Henry, D.B., Farrell, A.D., Schoeny, M.E., Tolan, P.H., Dymnicki, A.B., 2011. Influence of school-level variables on aggression and associated attitudes of middle school students. *Journal of school psychology* 49 (5), 481–503.
- Hoareau, C., Ganier, F., Querrec, R., Corre, F., Buche, C., Le Corre, F., Buche, C., 2013. Evolution of cognitive load when learning a procedure in a virtual environment for training. *6th International Cognitive Load Theory Conference*, pp. 130–132.
- Hsieh, H.-F., Shannon, S.E., 2005. Three approaches to qualitative content analysis. *Qualitative health research* 15 (9), 1277–1288.
- IJsselstein, W., de Kort, Y., Poels, K., 2013. The Game Experience Questionnaire.
- Ingersoll, R.M., Smith, T.M., 2003. The wrong solution to the teacher shortage. *Educational Leadership* 60 (8), 30–33.
- Ingersoll, R.M., Strong, M., 2011. The impact of induction and mentoring programs for beginning teachers: A critical review of the research. *Review of educational research* 81 (2), 201–233.
- Jaspers, M.W., 2009. A comparison of usability methods for testing interactive health technologies: methodological aspects and empirical evidence. *International journal of medical informatics* 78 (5), 340–353.
- Jemmali, C., Bunian, S., Mambretti, A., El-Nasr, M.S., 2018. Educational game design: an empirical study of the effects of narrative. *Proceedings of the 13th International Conference on the Foundations of Digital Games*. ACM, p. 34.
- Junod, R.E.V., DuPaul, G.J., Jitendra, A.K., Volpe, R.J., Cleary, K.S., 2006. Classroom observations of students with and without ADHD: Differences across types of engagement. *Journal of School Psychology* 44 (2), 87–104.
- Kao, C.-P., Tsai, C.-C., Shih, M., 2014. Development of a survey to measure self-efficacy and attitudes toward web-based professional development among elementary school teachers. *Journal of Educational Technology & Society* 17 (4), 302–315.
- Kazdin, A.E., 2008. *Parent management training: Treatment for oppositional, aggressive, and antisocial behavior in children and adolescents*. Oxford University Press.
- Kelleci, Ö., Aksoy, N.C., 2020. Using game-based virtual classroom simulation in teacher training: User experience research. *Simulation & Gaming*.
- Kervin, L., Ferry, B., Carrington, L., 2006. Classsim: preparing tomorrow's teachers for classroom reality. *Society for Information Technology & Teacher Education International Conference*. Association for the Advancement of Computing in Education (AACE), pp. 3204–3211.
- Killion, J., 2013. Meet the promise of content standards: Tapping technology to enhance professional learning. *Learning Forward*.
- Kökner, C., 2015. Modeling student behaviours in a virtual classroom with incorporation of social learning theory into belief-desire-intention model (yayınlanmam yüksek lisans tezi). *Orta Dou Teknik Üniversitesi/GATE*, Ankara.
- Kolb, A.Y., Kolb, D.A., 2009. *Experiential learning theory: A dynamic, holistic approach to management learning, education and development*. The SAGE handbook of management learning, education and development 42, 68.
- Kwon, J.H., Powell, J., Chalmers, A., 2013. How level of realism influences anxiety in virtual reality environments for a job interview. *International Journal of Human Computer Studies*.
- Le Corre, F., Hoareau, C., Ganier, F., Buche, C., Querrec, R., 2014. A pedagogical scenario language for virtual learning environment based on UML meta-model. *Proceedings of the 6th International Conference on Computer Supported Education-Volume 1*. SCITEPRESS-Science and Technology Publications, Lda, pp. 301–308.
- Lee, K.M., 2004. Presence, explicated. *Communication theory* 14 (1), 27–50.
- Lindsey, L., Berger, N., 2009. *Experiential approach to instruction*. *Instructional-Design Theories and Models, Volume III*. Routledge, pp. 129–154.
- Ludwick, D.A., Doucette, J., 2009. Adopting electronic medical records in primary care: lessons learned from health information systems implementation experience in seven countries. *International journal of medical informatics* 78 (1), 22–31.
- Lugrin, J.-L., Bossier, A.-G., Latoschik, M.E., Chollet, M., Glémarec, Y., Lugrin, B., 2019. Towards narrative-driven atmosphere for virtual classrooms. *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*, pp. 1–6.
- Lugrin, J.-L., Landeck, M., Latoschik, M.E., 2015. Avatar embodiment realism and virtual fitness training. *2015 IEEE Virtual Reality (VR)*. IEEE, pp. 225–226.
- Lugrin, J.-L., Latoschik, M.E., 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*.
- Malone, T.W., 1980. What makes things fun to learn? heuristics for designing instructional computer games. *Proceedings of the 3rd ACM SIGSMALL symposium and the first SIGPC symposium on Small systems*. ACM, pp. 162–169.
- Marion, N., Septeault, C., Boudinot, A., Querrec, R., 2007. Gaspar: Aviation management on an aircraft carrier using virtual reality. *Cyberworlds 2007 proceedings*, pp. 15–22.
- McLaughlin, T., Smith, D., Brown, I.A., 2010. A framework for evidence based visual style development for serious games. *Proceedings of the Fifth International Conference on the Foundations of Digital Games - FDG '10*, pp. 132–138.
- Merrill, M.D., 2009. *Instructional-design theories and models: Building a common knowledge base*. A Reigeluth CM & Carr-Chellman (ed.). Lawrence Erlbaum Associates 41–56.
- Mikropoulos, T.A., Natsis, A., 2011. Educational virtual environments: A ten-year review of empirical research (1999–2009). *Computers & Education* 56 (3), 769–780.
- Morgan, D.L., 1996. *Focus groups as qualitative research*, Vol. 16. Sage publications.
- Moskaliuk, J., Bertram, J., Cress, U., 2013. Training in virtual environments: putting theory into practice. *Ergonomics* 56 (2), 195–204.
- Nadolski, R.J., Hummel, H.G., Sloomaker, A., Van Der Veeg, W., 2012. Architectures for developing multiuser, immersive learning scenarios. *Simulation & Gaming* 43 (6), 825–852.
- Nagendran, A., Pillat, R., Kavanaugh, A., Welch, G., Hughes, C., 2014. A unified framework for individualized avatar-based interactions. *Presence: Teleoperators and Virtual Environments* 23 (2), 109–132.
- Nakayama, S., Jin, G., et al., 2015. Safety training enhancing outcomes through virtual environments. *Professional Safety* 60 (02), 34–38.
- Nakhla, B., Bevacqua, E., Said, B., Querrec, R., 2016. Cognitive architecture for embodied conversational agent: Application on virtual learning. *environment* 4973–4983.
- Nielsen, J., 1994. *Usability engineering*. Elsevier.
- Ochs, M., Mestre, D., De Montcheuil, G., Pergandi, J.-M., Saubesty, J., Lombardo, E., Francon, D., Blache, P., 2018. Training doctors social skills to break bad news: evaluation of the impact of virtual environment displays on the sense of presence. *Journal on Multimodal User Interfaces* 1–11.
- Oliver, R.M., Reschly, D.J., 2007. *Effective classroom management: Teacher preparation and professional development*. Washington, DC: National Comprehensive Center for Teacher Quality.
- Oliver, R.M., Reschly, D.J., 2010. Special education teacher preparation in classroom management: Implications for students with emotional and behavioral disorders. *Behavioral Disorders* 35 (3), 188–199.
- Querrec, R., Buche, C., Maffre, E., Chevaillier, P., 2004. Multiagents systems for virtual environment for training. application to fire-fighting. *International Journal of Computers and Applications* 1 (1), 25–34.
- Rayner, C., Fluck, A., 2014. Pre-service teachers' perceptions of simSchool as preparation for inclusive education: A pilot study. *Asia-Pacific Journal of Teacher Education* 42 (3), 37–41.
- Richey, R., Klein, J., Tracey, M., 2011. Chapter 2: General systems theory. *The Instructional Design Knowledge Base: Theory, Research and Practice* 11–29.
- Roldán-Álvarez, D., Martín, E., García-Herranz, M., Haya, P.A., 2016. Mind the gap: Impact on learnability of user interface design of authoring tools for teachers. *International Journal of Human-Computer Studies* 94, 18–34.
- Salas, E., Tannenbaum, S.I., Kraiger, K., Smith-Jentsch, K.A., 2012. The science of training and development in organizations: What matters in practice. *Psychological science in the public interest* 13 (2), 74–101.
- Schroeder, R., Axelsson, A.-S., 2006. *Avatars at work and play: Collaboration and interaction in shared virtual environments*, Vol. 34. Springer Science & Business Media.
- Schunk, D.H., 2008. Metacognition, self-regulation, and self-regulated learning: Research recommendations. *Educational psychology review* 20 (4), 463–467.
- Schussler, D., Frank, J., Lee, T.-K., Mahfouz, J., 2017. Using virtual role-play to enhance teacher candidate skills in responding to bullying. *Journal of Technology and Teacher Education* 25 (1), 91–120.
- Sherhoff, E., Frazier, S., Lisetti, C., Buche, C., Lunn, S., Brown, C., Delmarre, A., Chou, T., Gabbard, J., Morgan, E., 2018. Early career teacher professional development: Bridging simulation technology with evidence-based behavior management. *Journal of Technology and Teacher Education* 26 (2), 299–326.

- Shernoff, E., Frazier, S., Lisetti, C., Delmarre, A., Bibi, Z., Gabbard, J., 2020. Supporting the implementation of evidence-based behavior management practices through simulation: A mixed method study. *The Journal of Educational and Psychological Consultation*. (in press)
- Shernoff, E.S., Frazier, S.L., Marz-Lora, A.M., Lakind, D., Atkins, M.S., Jakobsons, L., Hamre, B.K., Bhaumik, D.K., Parker-Katz, M., Neal, J.W., Smylie, M.A., Patel, D.A., 2016. Expanding the role of school psychologists to support early career teachers: A mixed-method study. *School Psychology Review* 45 (2), 226–249.
- Shernoff, E.S., Lakind, D., Frazier, S.L., Jakobsons, L., 2015. Coaching early career teachers in urban elementary schools: A mixed-method study. *School Mental Health* 7 (1), 6–20.
- Shernoff, E.S., Maríñez-Lora, A.M., Frazier, S.L., Jakobsons, L.J., Atkins, M.S., Bonner, D., 2011. Teachers Supporting Teachers in Urban Schools: What Iterative Research Designs Can Teach Us. *School psychology review* 40 (4), 465–485.
- Shernoff, E.S., Von Schalscha, K., Gabbard, J.L., Delmarre, A., Frazier, S.L., Buche, C., Lisetti, C., 2020. Evaluating the usability and instructional design quality of interactive virtual training for teachers (ivt-t). *Educational Technology Research and Development* 1–28.
- Straub, C., Dieker, L., Hynes, M., Hughes, C., 2014. Using virtual rehearsal in the teachlive mixed reality classroom simulator to determine the effects on the performance of mathematics teachers. 2014 TeachLive National Research Project: Year 1 Findings 1–50.
- Tichon, J.G., 2007. Using presence to improve a virtual training environment. *CyberPsychology & Behavior* 10 (6), 781–788.
- Tracey, M.W., Hutchinson, A., Grzebyk, T.Q., 2014. Instructional designers as reflective practitioners: Developing professional identity through reflection. *Educational Technology Research and Development* 62 (3), 315–334.
- Wages, R., Grünvogel, S.M., Grützmacher, B., 2004. How realistic is realism? considerations on the aesthetics of computer games. In: Rauterberg, M. (Ed.), *Entertainment Computing – ICEC 2004*. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 216–225.
- Webster-Stratton, C., 2005. Teacher classroom management strategies questionnaire.
- Wenger, E., 1987. *Artificial intelligence and tutoring system: Computational and cognitive approaches to the communication of knowledge*. California: Morgan Kaufmann Publishers. Texto publicado na: *Pátio-revista pedagógica Editora Artes Médicas Sul Ano 1*, 19–21.
- Whyte, E.M., Smyth, J.M., Scherf, K.S., 2015. Designing serious game interventions for individuals with autism. *Journal of autism and developmental disorders* 45 (12), 3820–3831.
- Zibit, M., Gibson, D., 2005. simSchool: The game of teaching. *Innovate: Journal of Online Education* 1 (6).



Pr Cedric BUCHE is Professor in computer science at the National Engineering School of Brest (ENIB). He was the chief editor of the journal "International Journal of Virtual Reality" for 5 years. Pr BUCHE is currently in charge of a CNRS research team working on human system interaction.



Stacy Frazier, PhD, is a Professor in the Department of Psychology at Florida International University. She directs a federally-funded program of community-partnered and inter-agency research to strengthen youth service systems and support providers - including school personnel, afterschool staff, and early intervention specialists - to promote positive health and education trajectories for underserved youth.



Dr. Joseph L. Gabbard is an associate professor of Human Factors at Virginia Tech, where he directs the COGNITIVE Engineering for Novel Technologies (COGENT) Lab. He has MS and PhD degrees in Computer Science with a focus on HCI. His research program centers on human factors and human-computer interaction. Specifically examining perceptual and cognitive issues related to user experience and usability engineering of novel augmented and virtual environment user interfaces.



Alban Delmarre earned his PhD in Computer Science at Florida International University and is currently working at the Dan Marino Foundation. His research focuses on the use of virtual reality and virtual humans to benefit training and learning.



Dr. Shernoff earned her PhD in School Psychology at the University of Wisconsin-Madison and completed her post-doctoral fellowship at the University of Chicago. Prior to coming to Rutgers, she was a Research Assistant Professor of Psychology in Psychiatry (2006-2013) at the Institute for Juvenile Research at the University of Illinois at Chicago.



Dr. Christine Letitia Lisetti is an Associate Professor in the School of Computing and Information Sciences at Florida International University, and the director of the Virtual Interactive Social AGENTS (VISAGE) Laboratory. Her research focusses on human-computer interaction with socially affective interactive agents that have applications in training social skills, in cybertherapy, in healthcare and in medicine.