10-16-2020

Interactive Virtual Training: Implementation for Early Career Teachers to Practice Classroom Behavior Management

Alban Delamarre
adela177@fiu.edu

FLORIDA INTERNATIONAL UNIVERSITY

Miami, Florida

INTERACTIVE VIRTUAL TRAINING: IMPLEMENTATION FOR EARLY CAREER TEACHERS TO PRACTICE CLASSROOM BEHAVIOR MANAGEMENT

A dissertation submitted in partial fulfillment of the

requirements for the degree of

DOCTOR OF PHILOSOPHY

in

COMPUTER SCIENCE

by

Alban Delamarre

2020

To: Dean John L. Volakis
College of Engineering and Computing

This dissertation, written by Alban Delamarre, and entitled Interactive Virtual Training: Implementation for Early Career Teachers to Practice Classroom Behavior Management, having been approved in respect to style and intellectual content, is referred to you for judgment.

We have read this dissertation and recommend that it be approved.

Peter Clarke

Mark Finlayson

Jorge Riera

Monique Ross

Cédric Buche

Christine Lisetti, Major Professor

Date of Defense: October 16, 2020

The dissertation of Alban Delamarre is approved.

Dean John L. Volakis
College of Engineering and Computing

Andrés G. Gil
Vice President for Research and Economic Development
and Dean of the University Graduate School

Florida International University, 2020

ACKNOWLEDGMENTS

I would like to acknowledge my two advisors, Christine lisetti and Cédric Buche, who mentored me throughout the entire process. Over the years, Christine and Cédric were outstanding mentors who demonstrated concern for my professional and academic development. They facilitated avenues to enable me to develop as a scholar and researcher and made this work possible. Along these lines, I wish to thank the members of my dissertation committee: Monique Ross, Peter Clarke, Mark Finlayson, and Jorge Riera for their advice and support to conduct my research.

This dissertation represents a collaborative research project that involved many collaborators whom I thank: Dr. Elisa Shernoff, Dr. Stacy Frazier, and Dr. Gabbard.

This dissertation was partially funded by the US Department of Education and by the National Science Foundation. This dissertation was also financially supported in part through a Florida International University Graduate School Dissertation Year Fellowship.

I would like to address a special thanks to my parents Nathalie Beven and Denis Delamarre and my brothers Drs. Axel and Amaury Delamarre who supported and encouraged me along the way. Many thanks also to the VISAGE Lab team members, especially Stephanie Lunn for her unwavering friendship and encouragement and Santiago Bolivar for his never ending help and support. Finally, I want to thank Shelby Servais for her moral support, and the many discussions we had that motivated me to think deeper, consider new approaches, and ultimately finish my dissertation.

ABSTRACT OF THE DISSERTATION

INTERACTIVE VIRTUAL TRAINING: IMPLEMENTATION FOR EARLY CAREER TEACHERS TO PRACTICE CLASSROOM BEHAVIOR MANAGEMENT

by

Alban Delamarre

Florida International University, 2020

Miami, Florida

Professor Christine Lisetti, Major Professor

Teachers that are equipped with the skills to manage and prevent disruptive behaviors increase the potential for their students to achieve academically and socially. Student success increases when prevention strategies and effective classroom behavior management (CBM) are implemented in the classroom. However, teachers with less than 5 years of experience, early career teachers (ECTs), are often ill equipped to handle disruptive students. ECTs describe disruptive behavior as a major factor for stress given their limited training in CBM. As a result, disruptive behaviors are reported by ECTs as one of the main reasons for leaving the field.

Virtual training environments (VTEs) combined with advances in virtual social agents can support the training of CBM. Although VTEs for teachers already exist, requirements to guide future research and development of similar training systems have not been defined. We propose a set of six requirements for VTEs for teachers. Our requirements were established from a survey of the literature and from iterative lifecycle activities to build our own VTE for teachers. We present different evaluations of our VTE using methodologies and metrics we developed to assess whether all requirements were met. Our VTE simulates interactions with virtual animated students based on real classroom situations to help ECTs practice their CBM.

We enhanced our classroom simulator to further explore two aspects of our requirements: (1) interaction devices and (2) emotional virtual agents. Interactions devices were explored by comparing the effect of immersive technologies on users' experience (UX) such as presence, co-presence, engagement and believability. We adapted our VTE originally built for desktop computer, to be compatible with two immersive VR platforms. Results show that our VTE generates high levels of UX across all VR platforms. Furthermore, we enhanced our virtual students to display emotions using facial expressions as current studies do not address whether emotional virtual agents provide the same level of UX across different VR platforms. We assessed the effects of VR platforms and display of emotions on UX. Our analysis shows that facial expressions have greater impact when using a desktop computer. We propose future work on immersive VTEs using emotional virtual agents.

TABLE OF CONTENTS

CHAPTER PAGE

1. INTRODUCTION 1
1.1 Virtual Training Environments for Teachers 2
1.1.1 Requirements 4
1.1.2 Virtual Reality Platforms 5
1.1.3 Affective agents 7
1.2 Research Questions 8
1.3 Chapter descriptions 10

2. RELATED RESEARCH 12
2.1 Requirements for Virtual Training Environment for Teachers 12
2.1.1 Behavioral Fidelity 13
2.1.2 Environment Fidelity 19
2.1.3 Instructional Design 23
2.1.4 Autonomy 27
2.1.5 Interactivity 30
2.1.6 Scalability 34
2.2 VR Platforms for Virtual Training Environment for Teachers 36
2.2.1 Measures 37
2.2.2 VR Platforms Comparison 39
2.2.3 Usability and UX Evaluation of VR Platforms 41
2.3 Affective Non-Verbal Behaviors 46

3. CONTRIBUTION: IVT-T - INITIAL VERSION 50
3.1 Overview of IVT-T Context and Development Lifecycle 50
3.1.1 Classroom Behavior Management Strategies in IVT-T 50
3.1.2 IVT-T Initial Requirements 53
3.1.3 IVT-T Lifecycle Overview 54
3.2 Overview 57
3.3 Architecture 58
3.4 IVT-T's Graphics 61
3.4.1 3D Classrooms 61
3.4.2 3D Virtual Students 63
3.5 IVT-T Simulator 66
3.5.1 Vignette 66
3.5.2 3D Simulation 67
3.5.3 From vignette to 3D simulation 67
3.6 IVT-T User Interface 69
3.7 IVT-T Requirements' Evaluation 73
3.7.1 Behavioral Fidelity: Vignette scenarios 74
3.7.2 Environment Fidelity: 3D Classrooms and 3D Virtual Students 77

3.7.3 Interactivity: Usability and UX . . . 81
3.7.4 Instructional Design: Current Findings . . . 84
3.8 Discussion . . . 87

4. CONTRIBUTION: METIS - ADVANCED VR AND AFFECT VERSION 91
4.1 METIS' VR Platforms User Interface . . . 92
4.1.1 Differences between VR platforms . . . 94
4.1.2 Tutorial . . . 96
4.2 METIS Virtual students display of emotions . . . 97
4.2.1 Facial expressions animations . . . 97
4.2.2 Behavior Markup Language . . . 97
4.3 Evaluation of METIS: Usability and UX . . . 99
4.3.1 Methodology . . . 100
4.3.2 Results . . . 104
4.3.3 Discussion . . . 110
4.4 Evaluation of METIS: VR Platforms and Emotional Students . . . 117
4.4.1 Methodology . . . 118
4.4.2 Results . . . 123
4.4.3 Discussion . . . 129

5. FUTURE WORK . . . 134
5.1 Effect of facial expressions on VR . . . 134
5.2 Affective Autonomous Model for Virtual Students . . . 135
5.3 Immersive Interactions . . . 138
5.4 Vignette Authoring System . . . 140

6. CONCLUSION . . . 141

A. QUESTIONNAIRES . . . 143

BIBLIOGRAPHY . . . 149

VITA . . . 171

LIST OF FIGURES

FIGURE PAGE

3.1 Disruptive Behavior Evolution 52

3.2 IVT-T development lifecycle 55

3.3 IVT-T High-Level System Architecture 58

3.4 IVT-T Virtual Classrooms 62

3.5 IVT-T Virtual Students version 4 - 1^{st} grade students 64

3.6 Vignette to UML Activity Diagram 68

3.7 IVT-T Practice 70

3.8 IVT-T Reflection 71

3.9 IVT-T Feedback 72

3.10 IVT-T Storyline Summary 73

3.11 IVT-T Confidence Ratings 74

3.12 IVT-T Vignettes Logic and Realism 76

3.13 IVT-T Virtual Classrooms Realism 80

3.14 IVT-T Virtual Students Realism 80

3.15 IVT-T Usability: QUIS Results 84

3.16 IVT-T Practice Sessions Completion 88

4.1 METIS High-Level System Architecture 91

4.2 METIS UI: Student Speech Bubble 93

4.3 METIS UI: Decision 94

4.4 METIS CAVE and HMD 95

4.5 METIS Tutorial 96

4.6 METIS Facial Expressions 98

4.7 METIS Usability: QUIS Results 105

4.8 METIS UX: Technology Adoption Ratings. 107

4.9 METIS UX: Experience Consequence Ratings 109
4.10 METIS UX: Users' Emotions Ratings 109
4.11 METIS UX: Judgment Ratings 111
4.12 METIS VR Platforms and Emotional Student Experiment Procedure . . 121
4.13 METIS: Presence Ratings 124
4.14 METIS: Co-presence Ratings 125
4.15 METIS: Engagement Ratings 126
4.16 METIS: Believability Ratings 127
5.1 AIMER High-Level Architecture 139

LIST OF ACRONYMS

3B Breaking Bad Behaviors.

3D Three Dimensions.

ABC antecedent-behavior-consequence.

AUs Action Units.

BML Behavior Markup Language.

CAVE Cave Automatic Virtual Environment.

CBM Classroom Behavior Management.

CTA Concurrent Think-Aloud.

ECT Early Career Teachers.

HMD Head-Mounted Display.

ITS Intelligent Tutoring System.

IVT-T Interactive Virtual Training for Teachers.

MASCARET Multi-Agent Systems to simulate Collaborative, Adaptive and Realistic Environments for Training.

METIS Modeling Emotions for Training in Immersive Simulations.

QUIS Questionnaire for User Satisfaction.

SUS System Usability Scale.

UI User Interface.

UML Unified Modeling Language.

UX User Experience.

VR Virtual Reality.

VTE Virtual Training Environments.

WOz Wizard of Oz.

CHAPTER 1

INTRODUCTION

The use of virtual environments has grown exponentially in the past decades. Progress in computer graphics, visual display devices, interaction devices and the ability to integrate sounds with stereo effects has made possible the creation of Three Dimensions (3D) Virtual Training Environments (VTE) where users can observe, interact, practice, and learn. VTEs have been applied in a variety of domains, e.g. fire-fighter training, army soldier training, procedural training, safety training, and risk environment training [Querrec et al., 2004, Gerbaud et al., 2008, Nakhal et al., 2016, Nakayama et al., 2015, Barot et al., 2013].

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 by producing situations that require user intervention, which provides a hands-on experience.

VTEs have also been used for the development of social skills [Ochs et al., 2018, Kwon et al., 2013, Johnsen and Lok, 2008]. However, social skill VTEs require the use of interactive virtual humans able to exhibit social behaviors. Like virtual environments, the development of 3D realistic virtual human-like characters (also known as virtual agents, virtual humans, or avatars [1]) [Magnentat-Thalman and Thalmann, 2005] have observed a parallel growth. VTEs populated by virtual agents

[1]Because the two terms are often used interchangeably even though they have different meanings, we will now refer to virtual entities controlled by humans as avatars, and the terms "virtual agents" to refer to entities controlled by the system providing them some level of autonomy (as opposed to avatars).

are used for a variety of application such as the training of communication skills in high stress situations [Ochs et al., 2018], team collaboration [Robb et al., 2015], and in teaching context where classroom teachers can be confronted to disruptive virtual students [Dieker et al., 2015, Kervin et al., 2006, Gregory et al., 2013, Christensen et al., 2011, Lugrin et al., 2016].

1.1 Virtual Training Environments for Teachers

Disruptive behavior in the classroom is the main cause of stress for teachers, particularly for Early Career Teachers (ECT) (teachers with less than five years of experience) [Shernoff et al., 2011, Shernoff et al., 2016]. 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 teachers' training by tackling teachers' main issues: (1) teachers are unprepared for the realities of teaching [Grossman and McDonald, 2008]; (2) teachers have few opportunities to practice while receiving expertly tailored feedback about their classroom management [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 classroom 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, Oliver and Reschly, 2007]. Prevention and management of disruptive behaviors promotes student success by increasing the effectiveness of teacher instruction, and maximizing learning opportunities [Creemers, 1994, Crone and Teddlie, 1995, Oliver and Reschly, 2007]. Most of the time, teachers go through a trial-and-error approach in real classrooms to improve behavior management skills, resulting in teachers and students having a negative learning experience, an uncomfortable classroom climate, and strained interpersonal relationships [Henry et al., 2011a, Sabers et al., 1991]. Supporting the learning of Classroom Behavior Management (CBM) skills can mitigate negative classroom interactions, and can lead to facilitation of student learning [Creemers, 1994, Crone and Teddlie, 1995, Oliver and Reschly, 2007].

By simulating virtual classroom interactions VTEs allow teachers to experiment their CBM skills, and therefore provide a viable alternative for issues faced by teachers. VTEs designed to help teachers develop CBM skills have been used successfully [Zibit and Gibson, 2005, Kervin et al., 2006, Dieker et al., 2015, Gregory et al., 2013, Hayes et al., 2013b, Straub et al., 2014, Barmaki and Hughes, 2015, Tichon, 2007, Gupta et al., 2008]. VTEs offer a psychologically safe environment where a teacher's mistakes in managing the virtual students have no impact on real students. VTEs can also help systematically monitor teachers' training, and provide feedback on teachers' performance with the virtual students.

Building VTEs for learning and enhancing social skills, however presents many scientific challenges. In this dissertation, we aim to address the following: 1) identifying the functional, non-functional, and User Experience (UX) requirements needed for the design and implementation of effective training systems for teachers, as well as an appropriate system architecture to facilitate VTE implementation; 2) selecting the most efficient and effective interaction medium for the training of social skills (e.g. Desktop based vs. immersive technologies), given the hardware, software and human resources of the training program; and 3) designing and implementing realistic affective interactions between the virtual agents and the users.

We present the social skills VTE requirements established for building the Interactive Virtual Training for Teachers (IVT-T) system. IVT-T simulates disruptive student behaviors in a virtual classroom to provide practice experiences to teachers. Each practice session is followed by reflection opportunities on the teachers decisions to address the disruptive behaviors and teachers are given tailored feedback on their performance. Using the Modeling Emotions for Training in Immersive Simulations (METIS), an advanced cross-platform version of IVT-T that includes virtual students able to display facial expressions, we evaluated the effect of different Virtual Reality (VR) platforms and the effect of facial expressions on users.

Our goals are to:

1. **Implement the IVT-T system and its components**;
2. **Establish functional, UX, and graphical requirements for the creation of social skills VTE** based on our review of existing training systems for teachers and on our methodology while building IVT-T. We reveal important features that VTEs for the training of social skills must include in order to be efficient and effective;
3. **Integrate virtual agents able to display facial expressions into METIS and provide classroom simulations experience using immersive technologies or platforms** such as Head-Mounted Display (HMD) or Cave Automatic Virtual Environment (CAVE);
4. **Evaluate the usability of METIS across the platforms**; and
5. **Compare the effect of the three VR platforms (Desktop, HMD, CAVE) on teachers' perception of virtual students' emotions**;

1.1.1 Requirements

Requirements act as a contract between product owners, stakeholders and developers, and defines goals and metrics for the evaluation of the application. Using the appropriate requirements elicitation and analysis techniques prevents errors and lowers cost. Fixing errors in later stages of a system development is more expensive than during early stages [Davis, 1993]. Additionally, requirements ensure that the final product meets user needs as measured by the metrics specified in the requirements [Pandey and Pandey, 2012]. Establishing requirements is a crucial first step for the development of an effective software system.

Identifying a set of functional, non-functional, and UX requirements needed for the design and implementation of effective and engaging training systems for teachers presents many technical challenges, including: **1)** designing a system with a modular architecture so that it can be expanded upon; **2)** deciding upon the appropriate level of graphical realism of the 3D virtual environment and students, while weighing factors such as costs and performance; **3)** 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); **4)** finding the right evaluation methodologies to assess the numerous and varied aspects of the system and its User Interface (UI); **5)** integrating features to maintain users' engagement with the system by borrowing from game design elements; among other challenges.

1.1.2 Virtual Reality Platforms

VR platform refers to the type of devices used to show the virtual environment to users. With the development of immersive technologies, desktop setups are considered as a traditional VR platform. Immersive technologies refer to devices that provide inclusive, extensive, surrounding and vivid experiences to users [Slater and Wilbur, 1997]. Immersive technologies are becoming increasingly popular and accessible and are now being used to solve real world problems. The development of immersive technologies opens new opportunities for VTEs, which are now able to totally immerse users in the virtual environment and provide enhanced interactions as users are not constrained to a computer station.

With the growth of immersive VR, researchers have studied the effects of VR on users and on the potential for VR to increase learning [Mikropoulos and Natsis, 2011, Buttussi and Chittaro, 2018]. In general, studies show an increase in user engagement and sense of presence for immersive VR, such as HMDs or Cave Automatic Virtual Environment (CAVE, environment projected on walls surrounding the users), compared to traditional desktop applications [Buttussi and Chittaro, 2018, Ochs et al., 2018, Kim et al., 2014]. User engagement goes beyond user satisfaction, as the capacity of engaging users in a virtual environment is a crucial factor for e-learning activities within 3D-environments [Keller and Suzuki, 2004]. Similarly, the sense of presence, the feeling of "being there" in the virtual environment, indicates users' involvement and supports learning [Mikropoulos and Natsis, 2011].

HMDs are more accessible than ever before, but public perception may limit adoption of VR technologies for training purposes. Some users may think that HMDs are difficult to setup, uncomfortable to use for long periods of time, or simply not as practical as traditional training methods. Similarly, CAVE technology is expensive and not transportable. By developing VTEs that can be implemented across different VR platforms (hence cross-platform), it is possible to reach a wider audience and adapt the technology to user preference and proficiency. Users experiencing difficulty learning with the desktop VTEs, could be directed to HMD and CAVE platforms to better support their learning. However, the use of immersive VR for teacher training is still in its infancy and is yet to be explored [Lugrin et al., 2019]. Determining which VR platform yields the most efficient training would provide a great insight for the future development of training systems.

Whereas research comparing the effects of immersive technologies can provide important insights about their impact on users' experience (e.g. engagement, transfer of learning), current studies do not address how to design the UI to support comparisons across platforms. For effective comparisons, the UI designs must be adapted to the platform to provide comparable usability. Cross-platform VTE can accommodate users' preferences, proficiency, and platform availability, however, with every new technology, end-users acceptance is a critical factor for the success of new approaches. Users' perception of a system is impacted by the system's usability (i.e. ease of use, learn, and recall) [Nielsen, 1994, Hartson and Pyla, 2013]. Low usability can lead to the rejection of a system by users, no matter its potential [Napa et al., 2019]. Designing usable applications is a complex process and is still an ongoing debate in the computer science and in VR [Sutcliffe et al., 2019].

1.1.3 Affective agents

By taking advantages of progress on 3D realistic virtual human-like characters [Magnentat-Thalman and Thalmann, 2005], VTEs also benefit the domain of social skills training [Lisetti et al., 2013] for variety of domains such as the training of communication skills [Ochs et al., 2018], team collaboration [Robb et al., 2015] and teacher training [Dieker et al., 2015, Kervin et al., 2006, Gregory et al., 2013, Christensen et al., 2011, Lugrin et al., 2016].

The motivation to build virtual agents that can display emotions arose from psychology studies which demonstrated that emotions played a major role in social communications [Evans, 2002, Christianson and Loftus, 1991, Baron, 1987, Isen et al., 1987, Cialdini, 2009, Ekman, 2004]. Computer science researchers observed that emotional virtual agents were able to affect users in different context such as education [Lester et al., 2000], collaboration [Beale and Creed, 2009], and video games [Hamdy and King, 2017], among others. Researchers explored ways to use emotional virtual agents to generate desirable outcome such as learning [Beale and Creed, 2009].

Even though classrooms are highly emotional environments, the use of emotionally expressive virtual humans has not been studied in the context of virtual training for teachers. Previous research, however, demonstrated that emotional virtual agents have a positive effect on user's engagement [Pawel et al., 2009], motivation [Liew et al., 2017], and emotion contagion [Wu et al., 2014] on desktop computers. Exploring whether and how these effects transfer to immersive VR will inform the use of virtual agents displaying emotions in immersive simulations.

1.2 Research Questions

The objective of this thesis is to answer the following research questions (RQ) for VTEs for social skills training and VTEs designed for teacher training:

- **RQ1: What virtual reality and training specific requirements and system architecture should be considered in the creation of VTEs for K-12 teachers?** We will identify the functional, non-functional, and user experience requirements from our survey of existing training systems, and from our experience building IVT-T in collaboration with domain experts [Shernoff et al., 2018]; we will also recommend a computer architecture for VTEs based on our experience building virtual 3D environments;

- **RQ2: How to design and evaluate cross-platform VTEs' interaction design to provide comparable usability across three VR platforms (Desktop, HMD, CAVE)?** We describe our UI design methodology for the development of METIS, a virtual classroom simulator working on three VR platforms: (1) Desktop (**PC**); (2) Head-Mounted Display (**HMD**); and (3) Cave Automatic Virtual Environment (**CAVE**). We also recommend an approach to evaluate the cross-platform usability and aspects of UX such as technology adoption and cyber-sickness, and discuss insights for future development of cross-platform VTEs.
- **RQ3: What are the effects of a VTE experienced through three VR platforms (Desktop, HMD, CAVE) on teachers' UX in terms of engagement, presence, co-presence and their perception of the believability of the virtual students?** We will compare users' perception of virtual students' believability, engagement, and sense of presence and co-presence when using METIS across three different VR platforms (PC, HMD, CAVE).
- **RQ4: What are the effects of virtual students' display of emotions in a VTE on teacher's UX in terms of engagement, presence, co-presence and their perception of the believability of the virtual students?** We will enhance METIS by adding pre-scripted emotional virtual students able to display facial expressions and compare its effect on user' perception of agents' believability, engagement, and sense of presence and co-presence across three different VR platforms (PC, HMD, CAVE).
- **RQ5: For the three VR platforms (Desktop, HMD, CAVE) studied in RQ3, what are the effects of virtual students' display of emotions in a VTE on teacher's UX in terms of engagement, presence,**

co-presence and their perception of the believability of the virtual students? By comparing results from **RQ3** and **RQ4**, we will compare the effect of adding affective behaviors for each VR platform.

1.3 Chapter descriptions

This dissertation is structured as follows:

The first part of Chapter 2 presents a set of general requirements which must be considered and which will prove useful for researchers on VTEs for teachers. The requirements were established from our survey of the literature and our own lifecycle development. The second part of Chapter 2 provides a literature review of previous studies on cross-platform VTEs which compare the effect of different VR technologies on users. Existing work on the usability evaluation of desktop-based system and recent work on its implication for immersive VR systems are then presented. Finally, we underline the gap of the domain on emotional virtual agents when using immersive VR technologies.

Chapter 3 describe our approach for building IVT-T. We first establish the context in which IVT-T was developed and what were our initial requirements. We then describe IVT-T's development process and components. We present the different evaluations of IVT-T using methodologies and metrics we developed to assess whether all requirements have been met. We present the evaluation of the IVT-T UI usability. We also discuss current findings about the *soundness* of IVT-T instructional design.

In Chapter 4 we describe our UI design methodology for the development of METIS to be compatible with three technologies: PC, HMD, and CAVE. Additionally, we detail how METIS integrates virtual students that are able to display non-verbal behaviors and how we developed facial expressions for the students. Usability and other UX factors were evaluated for each with concurrent think-aloud protocol and semi-structured interviews. We present our results and discuss usability, technology adoption and cybersickness for future development of cross-platform VTEs.

In the last part of Chapter 4 we present the assessment of the effects of VR platforms and the display of facial expressions on *presence*, *co-presence*, *engagement* and *believability*. We present our results and their follow-up analysis to provide future research directions.

In Chapter 5 we describe METIS's next steps and clear the path for future research on affective virtual agents and immersive VR technologies. We conclude with a summary of our contributions in Chapter 6.

CHAPTER 2

RELATED RESEARCH

In this chapter, we first present how we extracted VTEs' requirements from the literature. We then review VTEs using immersive VR technologies and how usability and UX evaluation can be applied this type of technology. Finally, we describe existing research using emotional virtual agents.

2.1 Requirements for Virtual Training Environment for Teachers

Although existing VTEs for teachers have proven effective for some aspects of training [Zibit and Gibson, 2005, Kervin et al., 2006, Christensen et al., 2011, Dieker et al., 2015], a comprehensive set of requirements to guide the development of, and improve research on, VTEs for teachers does not exist, hence our first research questions (**RQ1**):

What virtual reality and training specific requirements and system architecture should be considered in the creation of VTEs for K-12 teachers?

As summarized in Table 2.1, 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 for teachers based on the project resources, and (2) to document how the most advanced VTEs for teachers have addressed these requirements. We conducted a survey of the literature by rendering explicit the implicit chosen requirements of existing VTEs for teachers research projects to address **RQ1**. We compiled our proposed set of requirement categories based on our analysis of existing VTEs for teachers, 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 requirement categories for VTEs for teachers, shown in Table 2.1, that we consider desirable (if not necessary) for the development of future VTEs for teachers, namely:

1. behavioral fidelity
2. environment fidelity
3. instructional design
4. autonomy
5. interactivity
6. scalability

2.1.1 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 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].

Realistic scenarios depicting the virtual behavior for the entities in VTEs are essential 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],

Virtual Training Systems for Teachers		ClassSim 	SimSchool 	Virtual PREX 	TeachLive 	3B 	SimInClass 	IVT-T 
Behavioral fidelity	Education experts' content	1 Scenario	Student model	12 Scenarios	Instructions	Instructions	16 Scenarios	9 Scenarios
	Affect	Still facial expressions	COVE Model	None	Facial expressions	None	None	None
Environment fidelity	Graphics	2D	2D	3D	3D	3D	3D	3D
	Student/Class	-	1 to 18	10	5	24	-	15
	Teacher representation	None	None	8 avatars (4 females)	None	1 avatar (1 white male)	None	None
	Animations	None	Still postures	~20 animations	Live motion capture	~25 animations	-	80 animations
	Audio	No	No	No	Instructors' voice	20 recordings	No	450 recordings
Instructional design	Adaptability to user	None	None	None	Instructor's initiative	Instructor's initiative	16 difficulty levels	3 difficulty levels
	Reflection	Thinking space	None	None	None	None	None	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
	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
Interactivity	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

Table 2.1: Comparison of VTEs for teachers based on the proposed requirements (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] and IVT-T [Delamarre et al., 2017]).

and relevant scenarios play a major role for motivating learning. Behavioral fidelity is also necessary for an efficient transfer of learning [Bossard et al., 2008, Dalgarno and Lee, 2009].

Three main approaches have been taken to generate virtual student behaviors:

- scenario-based: creating scenarios for each behavior at any given time that are then computerized and automatically controlled by the VTE,
- Wizard of Oz (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.

Scenarios-based approach

Education experts can create scenarios that map out how the virtual students act and react to teacher trainees' input during the virtual training simulations. Scenarios are then in turn programmed into the VTE for teachers to automatically control the virtual students. This approach was adopted for *ClassSim* [Kervin et al., 2006] to generate one scenario with 500 nodes.

VirtualPREX uses Second Life and is based on twelve role-play scenarios [Gregory et al., 2013]. A group of teacher trainees team up to play avatars during the scripted scenarios. Trainees alternate roles, for each scenario on teacher trainee is assigned 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, nine others are required to play the students.

For IVT-T, our education experts created nine scenarios (three scenarios for an off-task 1^{st} grader, three scenarios for an off-task 6^{th} grader, and three scenarios for an aggressive 6^{th} grader). We computerized these scenarios as described in Section 3.5, and evaluated them in terms of behavioral fidelity as described in Section 3.7.1. For each play of a scenario, teacher trainees are faced with making four to eight decisions. Overall, the scenarios contain around 50 different decisions.

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 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 for teachers, 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 evaluated 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., 2016] also makes use of a WOz to control virtual students' behaviors. During *3D* simulations, the WOz can adjust the level of disruption, [des/]activate a bad behavior out of six different behavior types, or 20 different dialogues 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.

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 20^9 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].

Behavioral Fidelity Validation

Regarding the validation of the behavioral realism requirements, we found that not all VTEs for teachers provided evaluation results. No evaluation for the behavioral realism of scenarios, or set of instructions provided to the WOz to generate these scenarios, were available for *simSchool*, and 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 [Badiee 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 [Gregory et al., 2013] 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 deter teachers from disregarding the system [Badiee and Kaufman, 2015] and prevent a 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 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 3.7.1) was conducted with a board of six retired teachers who, iteratively, rated four prototype versions of each scenario. The evaluation was conducted on multiple aspects, including the *logic and realism of students' behavior, actions, and dialogue according to their age* (1^{st} or 6^{th} grader) and *presenting problem* (inattention/hyperactivity or aggression/noncompliance), as well as the *teacher trainee response options*, and the *storyline engagement*.

We suggest that VTEs for teachers consider ensuring 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 for teachers behavioral fidelity.

2.1.2 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. Fidelity also concurs with the model of 3D VTE for learning proposed by Dalgarno et al. [Dalgarno and Lee, 2009] in which representational fidelity is a central characteristics to generate users' sense of presence, co-presence (i.e. the sense of "being there with someone") within the VTE [Lee, 2004, Bailenson et al., 2005]. Dalgarno et al. identify that textures, lightning, 3D models, frame per seconds, smooth view changes, spatial audio, and user representation are factors of the fidelity.

Graphics

There is an ongoing debate towards the realism of graphics in VTEs between researchers. Some posits that graphical fidelity could be detrimental for learning [Wages et al., 2004, Brenton et al., 2005] 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 were able to engendered higher levels of anxiety (as during actual interviews) as well as a higher sense of presence [Kwon et al., 2013]. Finally, a third approach advocates for a graphic realism adapted to the type of learning [McLaughlin et al., 2010]. Additionally, Bailenson et al. [Bailenson et al., 2005] showed that a mismatch between graphic 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 aligned for the development of VTEs for teachers.

In IVT-T, as the literature 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 also relied on iterative feedback from education experts to reach their desired level of realism.

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]. Felnhofer et al. [Felnhofer et al., 2019] observed an influence of attentive and emotionally responsive virtual agents on users' attention and stress, suggesting that classroom behavior responsive to teachers' actions can be beneficial for VTEs for teachers. In IVT-T classroom size were determined by education experts.

User representation

Regarding user representation in the VTE, only VirtualPREX [Gregory et al., 2013] and 3B [Lugrin et al., 2016] used avatars to represent the teacher. VirtualPREX created eight teacher avatars (four males and four females) for teacher trainees to choose from. Only one white male avatar is available for the 3B VTE for teachers. In order to not assume gender, race and ethnicity of the user, IVT-T do not integrates user representation. However, when prompted, virtual students will look at the user (the camera) as to indicate to users that they are in the virtual classroom.

Audio

Dalgarno et al. [Dalgarno and Lee, 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 where 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.

Environment Fidelity Validation

Some studies have looked into the evaluation of environment fidelity for driving simulations [Debattista et al., 2017] and urban planning [Drettakis et al., 2007]; however the metrics could not be applied to VTEs for teachers because they were domain specific (e.g. realism of a road or focusing on vegetation and crowd in front of a building). We created a validation scheme for virtual classrooms to include (*Physical Arrangement, Wall Decoration, Materials and Physical Appearance*), and for the virtual students in terms of *Face, Body, Clothing*, and *Hair*. A board of experienced teachers was recruited to evaluate IVT-T's environment realism (3.7.2).

Feedback on graphics can also be collected directly from teachers, as it was done with *VirtualPREX* [Gregory et al., 2013]. The 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 must be considered to ensure graphic realism.

Classroom size is an important aspect of VTEs for teachers, as the number of students can impact a teacher's stress level [Lugrin et al., 2016]. However, simulating a large number of students can be challenging, as more graphical rendering power would be required to smoothly run the VTE. Representing a realistic number of students increases the graphic requirement and can prevent teachers with slow computers from accessing the training system. Therefore, there is a trade-off between displaying a realistic number of students and the keeping the computational rendering power low.

Two audio integration approaches can be compared by observing two existing systems; *TeachLive* uses the voice of the WOz who modulates his/her voice to match the virtual students, and *3B* resorted to pre-recorded audios. Recorded human voices can be perceived as more understandable, expressive, and likeable than synthetic voices [Cabral et al., 2017]. Cabral et al. designed a questionnaire to evaluate virtual characters' voices and how they match the physical appearance of the virtual character [Cabral et al., 2017]. In IVT-T no questionnaire was used, however qualitative feedback on the children audio recordings was collected during interviews with participants.

2.1.3 Instructional Design

The Instructional Design requirement aims at optimizing the 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].

This Instructional Design requirement considers different aspects such as adaptability to user, user reflection and expert feedback.

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 VTEs for teachers relying on a WOz tele-operated approach such as *TeachLive* [Dieker et al., 2015] and *3B* [Lugrin et al., 2016], 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 for teachers using model-based approaches (*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 given on whether the students' parameters can be modified at run time to dynamically adapt the difficulty to the learner.

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.

Reflection

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 the pre-service teachers (Table 2.1). Authors noticed that pre-service teachers used this feature frequently, 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 four phases: Practice, Replay, Reflect, Feedback. During the Reflect phase, ECTs are specifically prompted to reflect on decisions they made during the practice phase. To assist ECTs in their reflections and help them remember specific decisions, IVT-T 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"*.

Feedback

Providing expert feedback is widely used among existing systems (see Table 2.1). Providing comments and feedback after practice help teachers acknowledge how they performed, adapt to new situations, and change their approach to improve future performances [Dieker et al., 2015, Lugrin et al., 2016]. 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]. Additionally, during traditional training, teachers have few opportunities to practice while receiving feedback [Denton and Hasbrouck, 2009, Shernoff et al., 2015]. Therefore, providing feedback is a necessary requirement for the creation of an effective training system. Feedback can be provided at runtime, Lugrin et al. [Lugrin et al., 2016] observed that giving audio feedback cues during the simulation did not affect the feeling of presence or the suspension of disbelief of users. Feedback can also be given after the simulation as quantitative or qualitative data. For instance, SimSchool [Christensen et al., 2011] 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 [Gregory et al., 2013], video of recorded sessions are used to give feedback to the teacher trainee.

Following the reflection phase, the IVT-T system provide ECTs with feedback from education experts based on the choices the user made to address the disruptive student behavior. In IVT-T expert feedback is pre-scripted for each choice in the scenarios and is provided as quantitative ($\sim$250 quantitative feedback per vignette) and qualitative feedback ($\sim$200 qualitative feedback per vignette).

Instructional Design 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 teachers being trained with the VTE for teachers compared to a control group following a traditional CBM training.

Before evaluating the efficiency of the training provided by the system which is a long process, it is possible to observe how teachers perceive and use the different aspect of the instructional design using formative evaluation. Feedback on the system can be collected using interviews or focus groups.

Access to IVT-T was given to a group of teachers from different schools and their interactions with the system were observed. Different elements of IVT-T's instructional design were evaluated and discussed through focus groups.

2.1.4 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 behavior management skills independently, or whether their training with the VTEs needs to be supervised. Technological devices required are also included in the autonomy requirement as these directly impact teachers' ability to train with the system.

Human resources

Existing classroom simulators using role playing or WOz setups increase the man power needed to run the system. In the case of *VirtualPREX* [Gregory et al., 2013], to train one teacher, the system requires ten actors to play (i.e. teleoperate) the ten virtual students present in the classroom, whereas *TeachLive* [Dieker et al., 2015] and *3B* [Lugrin et al., 2016] 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 for teachers 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 a study led by Badiee et al. [Badiee and Kaufman, 2015], an autonomous system using a model of behavior, such as *SimSchool* [Christensen et al., 2011] can result in users questioning the realism of the interactions and the situations presented.

The IVT-T system is completely autonomous: ECTs can practice, reflect and receive feedback without having to rely on an human actor. Moreover, IVT-T uses scenarios which were validated in terms of realism by experienced teachers (Section 3.7.1). Therefore, IVT-T provides a training platform with which ECTs can practice autonomously on realistic classroom situations.

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* [Lugrin et al., 2016] 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 is 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 immersive technologies such as HMDs or 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 highlights the need to conduct larger experiments to confirm their results. However, Lugrin et al. [Lugrin et al., 2016], observed that pre-service teachers responded positively to the use of immersive VR for the training of classroom management skills.

Autonomy Validation

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

Systems requiring personnel, specialized facilities, and equipment impose constraints on how frequently teachers can practice. For instance, in a study on the use of praise with *TeachLive*, some data points were missing because of teacher trainee absences. 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 teachers to go out of their way to access the training system.

2.1.5 Interactivity

Interactivity appears in all frameworks for VTE development [De Freitas and Oliver, 2006, Nadolski et al., 2012, Moskaliuk et al., 2013]. 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 include 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, easy to use and pleasant to interact with can avoid cognitive overload which in turn can negatively affect user learning experiences and willingness to engage with the system [Hartson and Pyla, 2018, Nielsen, 1994, Roldán-Álvarez et al., 2016].

Interactions

From the presented VTEs for teachers, we distinguish 3 types of interactions: menu choices, live text chat, and verbal and non-verbal communication. In ClassSim [Kervin et al., 2006], depending on the context of the situation, different options are presented to the trainees (e.g. reprimanding or ignoring students who speak without raising their hand; intervening or not when students are pushing one another to enter the classroom). SimSchool [Christensen et al., 2011] interactions are divided between two types of action: (1) attributing or adjusting tasks (e.g. asking to recite a poem or to work alone at one desk), or (2) addressing a virtual student behavior with an utterance. Users select from multiple options for each of these two types. In VirtualPREX [Gregory et al., 2013], the teachers playing the different roles communicate through the Second Life live text chat, and can select animations to be played by their avatars. Gregory et al. observed, however, that some users needed cheat sheets to remember how to control avatars. TeachLive [Dieker et al., 2015] and 3B [Lugrin et al., 2016], being controlled by human instructor, allow for verbal and non-verbal communications with the system.

Usability and UX

We recommend validating the usability and UX early in the development process [Chellali et al., 2016]. Identifying usability problems during the design lifecycle guides the overall development of a user-friendly system. Postponing this step to later development cycles could result in a system with poor usability and be disregarded by domain users. Moreover, trying to correct usability issues later in the development process can be costly and time consuming [Hartson and Pyla, 2018, Bowman et al., 2002b].

SimSchool [Christensen et al., 2011] and *VirtualPREX* [Gregory et al., 2013] 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 1 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. It is therefore beneficial for VTEs to present UIs that are usable and pleasant to interact with.

Interactivity Validation

Existing VTEs for teachers uses different types of interactions (menu choices, text chat, voice and gestures). User input based on voice and gestures allow for a natural interaction with the system, however their interpretation by a computer can be very challenging. This is why the two systems using voice and gestures interactions rely on WOz to control the system's response to users' input [Dieker et al., 2015, Lugrin et al., 2016]. A text chat approach faces similar challenges (interpreting users' text input to generate a corresponding system response). The VTE using text chat (*VirtualPREX*) also relies on human actors to control the system. Finally, menu choices are easier for computers to translate and respond too, but limit the options available to users. Increasing the number of options to cover a wide range of input can negatively impact the quality of the interaction with the system [Badiee and Kaufman, 2015]. For IVT-T, we relied on education experts to determine the type of interactions users would take to communicate with the system and the number of options to present to teachers (up to three options for each decision).

To evaluate the usability and UX we suggest a mixed-method approach, one which includes the collection of both quantitative (e.g. standardized rating scales and questionnaires) and qualitative (e.g. observations of user performance and semi-structured interviews) data. Using both types of data helps provide a more thorough and complete understanding of how the system is perceived by users than using each type of data separately [Creswell and Plano Clark, 2011].

The quantitative data can provide insights into the usability of the system and can be used to compare it to existing technologies. Quantitative data can provide an indication on user attitudes towards the system in terms of ease of learning, ease of use and technology adoption [Ludwick and Doucette, 2009]. Self-report questionnaires can help collect the 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] which are both standardized measures of the system usability. The QUIS provides data on the overall usability of the system, screen design and layout, terminology, learning, and system capabilities. Qualitative data can reveal potential usability problems or design flaws, which cannot be inferred from quantitative data. Methods such as the Concurrent Think-Aloud (CTA), during which users explain orally what they are thinking as they perform tasks which are recorded by a facilitator/observer next to the user [Cooke, 2010, Jaspers, 2009], 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 can be used to obtain a better understanding of user satisfaction with the graphics and/or instructional design elements [Morgan, 1996].

2.1.6 Scalability

Scalability, in this context, refers to the system ability to integrate scenarios or situations without requiring major changes in the implementation. This requirement coincides with Lugrin et al.'s *"extensibility"* requirement for 3B [Lugrin et al., 2016].

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, and the type of disruptive behavior. Presenting different training situations is desirable for VTEs as it generates abstraction which supports the transfer of learning [Bossard et al., 2008]. VTEs for teachers implementing the scalability requirement have the ability to integrate additional classroom situations without having to change the current implementation of the system and 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 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 to integrate variation in the classroom situation the users is facing. *SimSchool*, using a model-based system, can generate different reactions from the virtual students by changing the model parameters. Finally, for the *ClassSim* simulator, users interact with the system through a sequence of static web pages. To generate variability new static web pages must be created with their associated content (2D images, buttons).

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.

IVT-T uses the Multi-Agent Systems to simulate Collaborative, Adaptive and Realistic Environments for Training (MASCARET) [Querrec et al., 2004], a meta-model that provides a description of a virtual environment by interpreting Unified Modeling Language (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, provided that the corresponding 3D models and 3D animations are available.

In this section we presented the requirements we identified as beneficial for the development of VTEs for teacher training (**RQ1**). The requirements 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, Nadolski et al., 2012, Moskaliuk et al., 2013]; 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];

2.2 VR Platforms for Virtual Training Environment for Teachers

Existing classroom simulators took different approaches for the choice of the interaction device or platform (Table 2.1). ClassSim [Kervin et al., 2006], SimSchool [Rayner and Fluck, 2014] and VirtualPREX [Gregory et al., 2013] resorted to a desktop approach (mouse, keyboard and monitor). TeachLive [Dieker et al., 2015] presents teachers with a life-size classroom projected on the wall and 3B [Lugrin et al., 2016] resorts to a HMD. No justification is provided regarding the choice of a particular platform which suggests that little to no research has been done on identifying the best technology to train teachers. Why invest into immersive technologies (HMD, CAVE) if a traditional desktop approach provides similar results?

In this section we present existing studies on the comparisons of VR platforms. But first, we define and present measures generally considered for VR platforms comparisons.

2.2.1 Measures

Presence and Co-presence Presence has been defined by Steuer as *"the extent to which one feels present in the mediated environment, rather than in the immediate physical environment"* [Steuer, 1992]. Building on this definition and on previous work, Lee distinguished three types of presence: Physical, Social and Self presence [Lee, 2004]. Physical presence refers to the sense of "being there" in a virtual environment, in other words, it is the sense of being in the virtual place rather than in the physical place where one's body is located [Witmer and Singer, 1998, Slater and Steed, 2000]. Social presence, or Co-presence, is a psychological state where the virtual social actors are perceived as real social actors, i.e. co-presence occurs when a user does not feel that the virtual agents (autonomous virtual entities) that he/she is interacting with are artificial [Lee, 2004, Bailenson et al., 2005]. Self-presence refers to the feeling of identity construction inside the virtual environment [Lee, 2004]. The sense of presence is an important factor to consider for VTEs, as it has been positively associated with learning as an indicator of user involvement in a task [Mikropoulos and Natsis, 2011].

Given the social context of this study, we focused our observation on the effect of different VR platforms and of facial expressions on users' perceived presence and co-presence. Different factors can impact the feeling of presence: *(a) - Realism:* Realism of virtual environments is positively correlated with presence [Witmer and Singer, 1998]; *(b) - Quality:* Quality includes realism, fluidity, and the ability a VTE has to create interactions with users [Hendrix and Barfield, 1996]; *(c) - Ease of use:* VTEs that are easy to interact with have a positive effect on presence [Billinghurst and Weghorst, 1995]; *(d) - Control:* The users' sense of control correlates with their sense of presence [Witmer and Singer, 1998]; *(e) - Co-presence:* The ability to interact with others (virtual or real humans) and how they react also impact presence [Heeter, 1992]; *(f) - Exposure:* The duration of interaction with the virtual environment also plays a role in users' perceived presence. For instance, with a HMD, interactions longer that 15 minutes can have a negative effect on presence. In fact, exposure length is negatively correlated with presence [Witmer and Singer, 1998]. In a social skills training context, where social interactions are at the center of the training, the sense of co-presence (i.e. "being there with someone" in the VTE) is an important factor to consider [Ochs et al., 2018].

Engagement Engagement is defined as *"a value of user-experience that is dependent on numerous dimensions, comprising aesthetic appeal, novelty, usability of the system, the ability of the user to attend to and become involved in the experience and the user's overall evaluation of the salience of the experience"* [O'Brien and Toms, 2008]. Engagement is an aspect of UX which involves more than solely user satisfaction, as engaging users is a compelling factor for e-learning activities with 3D virtual environments [Mount et al., 2009, Keller and Suzuki, 2004].

Believability Bates defined believable characters not as an honest and reliable entity but as one that provides the "illusion of life" suspending disbelief in users [Bates et al., 1994]. As the visual aspect of a character is the most important features to generate realism [Togelius et al., 2013], believability on the other hand relies mainly on actions or behaviors displayed by the virtual character. A famous example given by Loyall highlights the differences between those two concepts [Loyall, 1997]: considering the Flying Carpet in the Aladdin Disney movie, Loyall showed that even without a mouth, eyes, or even a face, the Flying Carpet is absolutely not humanly realistic. However, the Flying Carpet demonstrates a personality with goals, motivations, and emotions. Studies focused on behavior as a communication medium displaying virtual agents' internal states using facial expressions [Malatesta et al., 2009], gaze [Poel et al., 2009], gestures [Corradini, 2004], or a combination of different means [Bevacqua et al., 2007]. However, what precisely constitute believability of characters vary among researchers. Hamdy and King [Hamdy and King, 2017] compiled a table of believability requirements from the literature. Affect and social relationships seems to be one of the most important aspects considered by researchers [Mateas, 1999, Loyall, 1997, Gomes et al., 2013, Lee and Heeter, 2015, Bogdanovych et al., 2016]. Another approach towards believability specifies that realistic, complex and highly intelligent behaviors are not necessary as long as the virtual agents' behaviors matches users' expectations in terms of personality and emotions [Dautenhahn, 1998]. Virtual agents also generate believability if they are coherent in their reactions and act consistently in similar kinds of situations [Ortony, 2003]. Believability is an important factor to consider as researchers argued that representational fidelity, in terms of graphics and behaviors, is necessary to achieve a highest transfer of learning [Dalgarno and Lee, 2009, Bossard et al., 2008].

2.2.2 VR Platforms Comparison

Many studies have compared the use of different VR platforms. However, these comparisons mostly focus on spatial orientation and navigation [Bowman et al., 2002a, Santos et al., 2009], data and object visualization [Mizell et al., 2002, Zielasko et al., 2016], procedure learning and memorization [Hirose et al., 2009, Buttussi and Chittaro, 2018], therapies and phobias [Juan and Pérez, 2009], and on the symptoms generated by the different VR technologies [Sharples et al., 2008].

Cummings and Bailenson conducted a survey of the literature to observe the impact of immersion on physical presence [Cummings and Bailenson, 2016]. Immersion, here, is considered as technological characteristics of a device as defined by

Slater and Wilbur: *"Immersion is a description of a technology, and describes the extent to which the computer displays are capable of delivering an inclusive, extensive, surrounding and vivid illusion of reality to the senses of a human participant"* [Slater and Wilbur, 1997]. Cummings and Bailenson reviewed 32 studies comparing low versus high immersion during various tasks such as navigation, search tasks or therapy treatment. The analysis revealed that generally, more immersive setups resulted in a stronger sense of presence. Moreover, they observed that features like stereoscopic visuals or wider fields of view had a much more significant impact on presence than graphics or auditory stimuli. Cummings and Bailenson however noted that results obtained by the studies considered could only be applied for physical presence, and could not be carried over to co-presence. Cummings and Bailenson argued that situation models constructed by users to experience co-presence must be built on communication channels rather than on spatial cues. None of the studies included in the survey were directed towards the training of social skills, and virtual humans in these virtual environments did not display emotions.

Few studies aimed at training social skills using virtual humans compared the effect of different VR platforms. Zanbaka et al. [Zanbaka et al., 2007] compared the social inhibition of completing a complex task when being observed by a real human, by a virtual life-size human projected on the wall and by a virtual human viewed through a HMD. Their results showed that inhibition, can be felt from the presence of a virtual human. Johnsen et al. [Johnsen and Lok, 2008] compared a HMD with a life-size projection on the wall to train social skills to medical students to prepare them for interacting with patients. Results showed that the HMD decreased medical students' ability to self-evaluate their empathy. However, Johnsen et al. posit that the nature and novelty of the HMD, in 2008, may have distracted medical students from the virtual patient and biased the results.

A recent study realized by Ochs et al. [Ochs et al., 2018] on the training of communication skills for medical experts - where users impersonate a doctor to deliver bad news to a virtual patient - compared a 3D-desktop with a HMD and a CAVE. The study included two types of participants: medical experts and participants with no medical experience (naive participants). Results showed that the CAVE and HMD setups yield more presence and co-presence than a desktop setup, and that experts tend to be more involved in the interactions with the virtual patient that naive participants. Additionally, whereas the system includes affective non-verbal behaviors for the virtual human, no indications are provided on which emotions were portrayed and how they were portrayed. However, as the study focus was on the difference of perception between medical experts and naive participants, the interaction between VR platforms and affective non-verbal behaviors was not explored.

The use of immersive VR for teacher training is still in its infancy and is yet to be explored [Lugrin et al., 2019]. Researchers emphasize the need for more studies to offer a deeper insight in the use of virtual human with immersive VR [Ochs et al., 2018, Lugrin et al., 2016, Lugrin et al., 2019].

2.2.3 Usability and UX Evaluation of VR Platforms

When comparing the effectiveness of different VR platforms, the usability of the system for each platform should be carefully considered. With every new technology, end-users acceptance is a critical factor for the success of new approaches. Usability plays a major role in how any system is perceived and adopted by users [Hartson and Pyla, 2013]. To ensure sound comparisons and effectiveness across all platforms, the interaction design must provide comparable usability.

Existing studies show greater presence and engagement with immersive platforms compared to desktop setups [Buttussi and Chittaro, 2018, Zanbaka et al., 2005, Ochs et al., 2018, Kim et al., 2014]. However, usability evaluations of the different platform prototypes were not conducted for these studies. Instead, oral instructions were given to participants on how to interact with the system. In [Buttussi and Chittaro, 2018], a special level was created for the experimenter to explain the controls as participants learn how to operate the system. Ensuring participants are able to interact with the system can be time consuming, especially for larger studies. In [Lugrin et al., 2013], authors compared players' performances of a desktop setup and a CAVE setup for a first person shooter video game. During the training phase, participants familiarized themselves with the system while being assisted by the experimenter, Lugrin et al. specified that: *"great care was taken neither to disclose how users may actually maximize their scores under both settings, nor to demonstrate the use of immersive gaming by an experienced user"*[Lugrin et al., 2013]. This indicates that, in some cases, information given to participants could introduce bias during the training phase. Zaidi et al. [Zaidi et al., 2019] compared two approaches to deliver instructions: oral instructions given by the experimenter vs a tutorial level where users learn to interact by themselves. Zaidi et al. witnessed greater usability for the tutorial approach compared to the oral instructions approach [Zaidi et al., 2019]. Increasing the usability of VTEs can increase the independence of users and reduce the need for verbally coaching participants on how to use the system.

Focusing on usability early on in the development process can ensure interactions are adapted with the end-user in mind and meet requirements [Bowman et al., 2002b]. Several approaches to guide the design of virtual environment interactions have been proposed [Chen and Bowman, 2009, Blom and Beckhaus, 2014, Sutcliffe and Gault, 2004].

Chen and Bowman [Chen and Bowman, 2009] consider an architecture based on three levels: application, domain, and generic interactions. Chen and Bowman distinguish three main types of interactions using virtual environments: viewpoint motion control (navigation), selection and manipulation [Bowman, 1998]. Interactions are either device-based, actions in the virtual environments are carried out using remote controllers, or human-based where the body acts as the controller. Human-based controllers require tracking of body movements so they can be represented in the virtual environment. Blom and Beckhaus [Blom and Beckhaus, 2014] argue for an approach with dynamic components (i.e. changing over time) and interactions generating more engagement. However, these approaches failed to provide indications on when and how to apply the proposed interactions when designing a virtual environment for a specific domain [Sutcliffe et al., 2019].

Sutcliffe et al. [Sutcliffe and Gault, 2004] basing their work on Human-Computer Interaction knowledge, proposed 12 design heuristics for VR applications. These heuristics focus on creating interactions as close as possible to the' real world, providing clear feedback for user actions, and helping navigation and features exploration. However, in a recent study, Sutcliffe et al. [Sutcliffe et al., 2019] commented on the complexity of applying existing VR design frameworks [Chen and Bowman, 2009, Blom and Beckhaus, 2014] and heuristics [Sutcliffe and Gault, 2004]. There is a trade-off between creating interactions that are usable and efficient and creating interactions that are realistic and immersive. Sutcliffe et al. argue that this trade-off can be settled by the domain requirements [Sutcliffe et al., 2019].

Many methods exist to evaluate the usability of interactive systems, however these methods have limitations when applied to virtual environment application where interactions are different from traditional desktop user interfaces [Bowman et al., 2002b]. Bowman et al. [Bowman et al., 2002b] distinguish four types of issues when applying usability evaluation methods to virtual environments: Physical environment issues, evaluator issues, user issues, and issues related to the type of usability evaluation. For example, if presence is being evaluated, the evaluator cannot intervene or be seen during the interaction which would break the feeling of presence and bias the results. However, participants with no experience with VR may experience difficulties interacting with the systems and require the evaluator to intervene. Additionally, some usability evaluation methods such as the CTA protocol, where users talk out loud about their actions and thoughts while they interact with the system also break the feeling of presence. Bowman et al. proposed a classification of usability evaluation methods for virtual environments based on three characteristics: context of evaluation, user involvement and type of results. For example, to evaluate the usability of a specific application with users, authors recommend a formative evaluation, a formal summative evaluation and post-hoc questionnaires to collect quantitative data, post-hoc questionnaires and interviews to collect qualitative data, and post-hoc interviews to collect qualitative data.

Several studies have explored the usability of cross-platform systems. Cao et al. [Cao et al., 2019] used the walk-through method, proposed by Sutcliffe et al. [Sutcliffe and Kaur, 2000], to evaluate the usability of a lunar exploration serious game played on both a traditional desktop and a HMD. The walkthrough approach consist of evaluating the user interface by stepping through common tasks to performed within the system. The interface capabilities were evaluated as the tasks are being performed.

None of the studies mentioned above used a the technique of the CTA to evaluate the usability of their cross-platform system. As specified by Bowman et al. [Bowman et al., 2002b], CTA for virtual environment can be a problem depending on the context of evaluation. A few studies used CTA with immersive VR systems, however the CTA was not used to evaluate usability but to measure conceptual learning [Roussou et al., 2006], observe conclusion drawn by users when observing data with a HMD [Millais et al., 2018], and examine users' feeling about fingerless hands representation [Schwind et al., 2017]. Napa et al. [Napa et al., 2019] used CTA to evaluate and compare the usability of two VR applications for Cardiac Surgery Case Planning using a HMD. CTA helped authors identify strengths and weaknesses of both applications. Additionally, CTA helped identify which features caused the most frustration for participants, however authors did not evaluated the usability of the same application on other platform.

The feedback collected using CTA reflect users first impression of their interaction with the system [Birns et al., 2002]. CTA allows evaluators to identify positive and negative aspects perceived by users testing the application, but the data is only qualitative. Basing their work on existing questionnaires, Tcha-Tokey et al. [Tcha-Tokey et al., 2016] developed a questionnaire for immersive virtual environments to help gather quantitative data for a variety of aspects of the UX. In addition to provide well accepted metrics such as presence, immersion, and engagement, this questionnaire rates the adoption of the technology by users, the attractiveness of the system, the cybersickness generated, and the emotional experience of the virtual environment.

The increased accessibility to immersive VR technology, brought attention on how to design usable system for this type of technology, however this topic is still ongoing research [Sutcliffe et al., 2019, Cao et al., 2019], hence our second research questions (**RQ2**):

How to design and evaluate cross-platform VTEs' interaction design to provide comparable usability across three VR platforms (Desktop, HMD, CAVE)?

2.3 Affective Non-Verbal Behaviors

The motivation behind the creation of virtual humans able to display emotions relies on psychological findings underlining the effect emotions have on attention [Evans, 2002], memory [Christianson and Loftus, 1991], judgment and decision making [Baron, 1987], creative problem solving [Isen et al., 1987], and persuasion [Cialdini, 2009]. Emotional expression also plays a major role in social communication [Ekman, 2004]. By creating embodied agents able to effectively and naturally communicate with users, researchers can aim to produce desirable or beneficial outcome such as transfer of learning [Beale and Creed, 2009]. Representational fidelity, in terms of graphics and behaviors, have been argued as necessary to achieve the highest transfer of learning [Dalgarno and Lee, 2009, Bossard et al., 2008]. Behavior fidelity is defined as *"the consistency of the objects behaviors, including the way that they respond to user actions and their autonomous (or modeled) behaviors"* [Dalgarno and Lee, 2009].

The definition provided by Dalgarno et al. [Dalgarno and Lee, 2009] can also be applied to virtual humans under the term believability. A believable character is one who gives the illusion of being alive, who can perform actions that make sense, and about whom users are able to suspend their disbelief [Mateas, 1999]. Research indicates that highly intelligent, realistic and complex behaviors are not necessarily required to achieve believability, as long as the virtual human displays behaviors which match users' expectations in terms of personality and emotions [Dautenhahn, 1998], are coherent in their reactions, and act consistently in similar kinds of situations [Ortony, 2003]. Moreover, the appearance of the virtual human has been shown to be one of many parameters impacting believability [Loyall, 1997]. Portraying affect and modeling social relationships are necessary features in order to allow the suspension of disbelief from users [Mateas, 1999, Loyall, 1997, Gomes et al., 2013, Bogdanovych et al., 2016]. Researchers of virtual humans aim to provide virtual humans with the capability to exhibit emotions through speech and non-verbal behaviors (body languages, facial expressions).

A survey of the literature conducted by Beale et al. [Beale and Creed, 2009] observed that few studies reported a negative impact of the display of emotion on the interaction with the system (Usability [Bartneck, 2003]) and on users (Enjoyment, persuasion and trust [Fabri et al., 2005, Berry et al., 2005]). The majority of the studies surveyed by Beale et al. [Beale and Creed, 2009] either indicate no impact or a positive effect of the display of emotions on the interactions with the system.

More recent studies, however, demonstrated the effect of virtual humans on users for different factors. For instance, Zanbaka et al. [Zanbaka et al., 2007] showed that social inhibition would be experienced by participants with both real and virtual humans. Pan et al. [Pan et al., 2011] observed increased level of psychological stress when they confronted their participants to moral dilemmas involving virtual humans. Others studies have showed the positive impact of virtual humans displaying emotions on engagement [Pawel et al., 2009], and motivation [Liew et al., 2017].

Emotional virtual humans are also used in the educational domain as a pedagogical agent [Lester et al., 2000, Moridis and Economides, 2012] using empathy to alter learners' emotional state (e.g. avoid frustration, provide encouragement). Of the existing classroom simulators [Kervin et al., 2006, Christensen et al., 2011, Gregory et al., 2011, Dieker et al., 2015, Lugrin et al., 2016], few resort to students displaying emotions, and little detail is given on how and when the emotions are displayed [Kervin et al., 2006, Dieker et al., 2015].

Few studies explored the use of immersive VR platforms and emotional virtual humans. Harjunen et al. [Harjunen et al., 2018] used a HMD and a haptic glove to observe the effects of facial expressions and touch of a virtual human on persuasion. The study showed that when the agents touched or smiled to participants, they were more likely to accept unfair offers. However, since only a HMD was used, the interaction between facial expressions and VR platform were not tested.

Ravenet et al. [Ravenet et al., 2016] compared the perception of attitudes of virtual humans in a conversational group with two different VR platforms (Desktop and CAVE). Conversational groups were composed of five members (four virtual agents and the user) placed in a circle. The user was tasked with determining the attitudes of two closest agents (left and right) based on their non-verbal behaviors, including turn-taking behaviors, gestures or interpersonal distance. Participants were able to correctly identify the attitudes in both setups, indicating that non-verbal behaviors from virtual humans are perceived similarly in non-immersive and in immersive setup.

However, the goal of this study was focused on user's recognition of attitudes and did not discuss the impact that these attitudes have on the users. Therefore, the impact on users of the VR platforms combined with the non-verbal behaviors was not explored. Moreover, users were only observers of the attitudes of the virtual agents. Users did not participate or interact with the conversational group, and the effects of the interaction between users and virtual agents for each VR platform were not observed.

Therefore, to our knowledge, no studies explored or compared the relationship between display technology (e.g. PC, HMD, CAVE) and non-verbal behaviors such as facial expressions exhibited by virtual humans. Given the trade-offs in cost, development, and resources for both the immersive technologies and emotional virtual agents, determining their effect on UX and ultimately learning, will support the design of optimal VTEs for social skills training.

To establish the interaction between display technology and emotional virtual agents we will first compare UX generated by the HMD and the CAVE compared to a PC with METIS (**RQ3**).

What are the effects of a VTE experienced through three VR platforms (Desktop, HMD, CAVE) on teachers' UX in terms of engagement, presence, co-presence and their perception of the believability of the virtual students?

Secondly, we will observe for all platforms the impact of facial expressions on UX (**RQ4**).

What are the effects of virtual students' display of emotions in a VTE on teacher's UX in terms of engagement, presence, co-presence and their perception of the believability of the virtual students?

Finally, we will compare the impact generated by the addition of facial expressions on UX between the VR platforms (**RQ5**).

For the three VR platforms (Desktop, HMD, CAVE) studied in RQ3, what are the effects of virtual students' display of emotions in a VTE on teacher's UX in terms of engagement, presence, co-presence and their perception of the believability of the virtual students?"

CHAPTER 3

CONTRIBUTION: IVT-T - INITIAL VERSION

To describe our approach at building IVT-T we first provide the context of IVT-T and how we conducted the development of the system. We then present the architecture and detail the components of IVT-T. Finally, we present the results of the evaluation of IVT-T for the proposed requirements for VTEs we established to address **RQ1**. The system described in this chapter was presented in published work [Delamarre et al., 2017, Delamarre et al., 2019b].

3.1 Overview of IVT-T Context and Development Lifecycle

In this section we present the theoretical background behind the development of IVT-T behavior representation. We also describe the first set of requirement describing the expectations of IVT-T's educational experts. We then detail IVT-T's development lifecycles.

3.1.1 Classroom Behavior Management Strategies in IVT-T

Although IVT-T can be useful to any teacher interested in improving their CBM, the goal of IVT-T is to provide ECTs with a realistic classroom teaching experience. IVT-T offers a low-stakes training environment and maximizes active learning opportunities. The difficulty of managing classroom behavior is exacerbated for ECTs who already receive limited mentoring in behavior management [**?**], 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 or feedback, which can be costly to teachers and their students [Henry et al., 2011b, 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 train to identify them during their practice with IVT-T:

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

In an *ABC cycle* [Kazdin, 2008], *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 facilitates student learning. *Proactive monitoring* consists of identifying early cues of disruptive behaviors. *Effective redirection* involves efficient, early and private redirection, combined with a consequence hierarchy as well as praises for student's compliance.

As shown in Figure 3.1, IVT-T scenarios were written 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 a student's attention, compliance, and engagement [Evertson and Weinstein, 2013, Junod et al., 2006, Kazdin, 2008]. Action consequences include the following options: *praise, ignore, redirect, use of proximity, instructions, empathy,* and *if/then statements.*

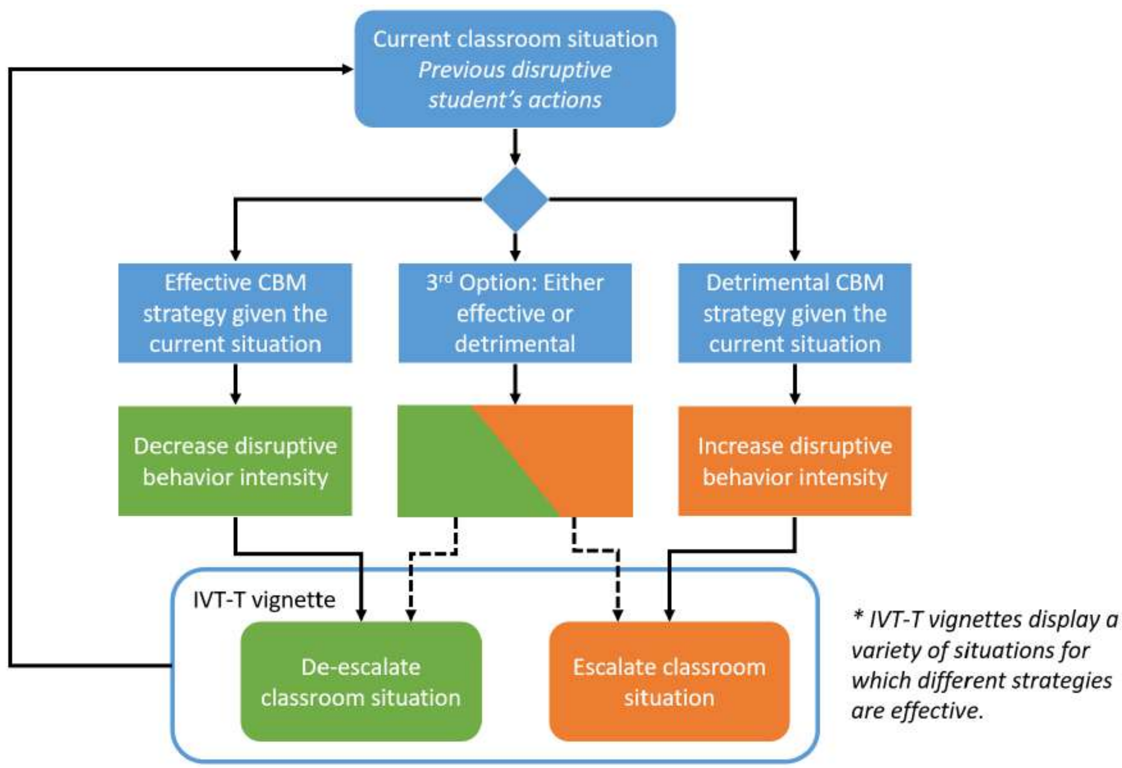


Figure 3.1: Disruptive Behavior Evolution. Evolution of the disruptive behaviors based on vignettes' context and on strategies selected by the trainee

Depending on the student behavior and on the situation, resorting to effective strategies deescalate the disruptive behaviors (e.g., the student becomes more engaged and compliant), whereas detrimental strategies escalate the situation (e.g., the 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).

3.1.2 IVT-T Initial Requirements

The education experts on the IVT-T 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 low income ECTs at any time, emphasizing the importance of online access and low-technology requirements (laptop or desktop computer).

3.1.3 IVT-T Lifecycle Overview

Our main research objectives for IVT-T were to:

1. identify what are the main *requirements* for building effective, usable, and enjoyable VTEs for teachers;
2. 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;
3. assess IVT-T *fidelity* in terms of whether it is used as intended; and
4. assess IVT-T *feasibility* in terms of transfer of knowledge and skills from the virtual classroom to ECTs live classrooms.

In this chapter we discuss how we reached our first three objectives which led us to build and validate IVT-T's realism, usability, and usage.

Given the 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 [Hartson and Pyla, 2018]. The IVT-T development lifecycle is depicted in Figure 3.2, 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.

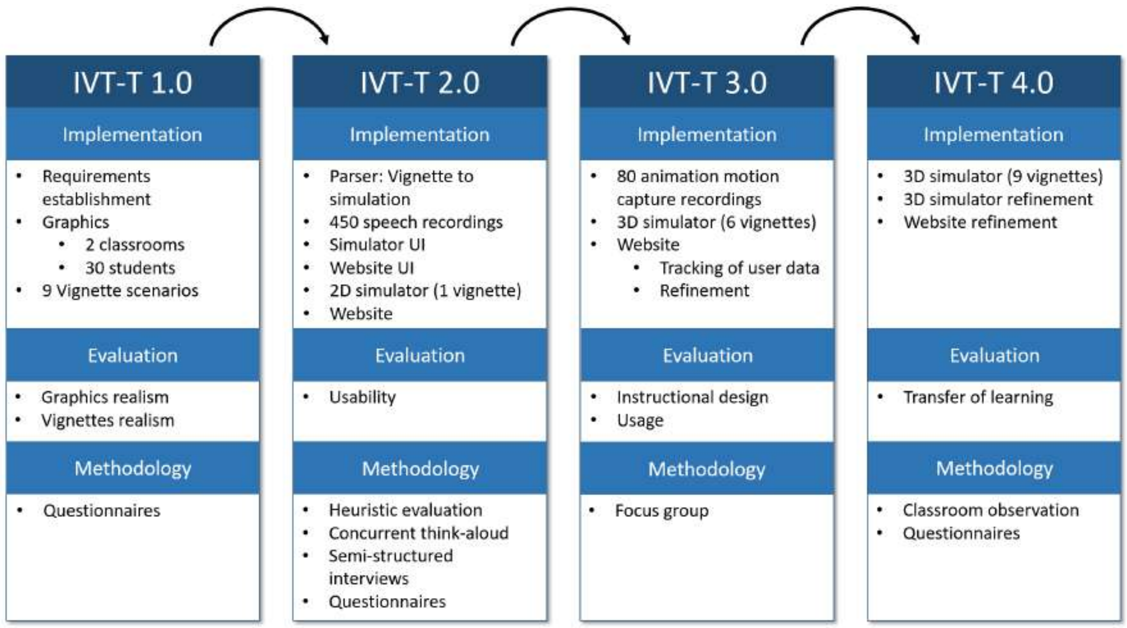


Figure 3.2: IVT-T development lifecycle

During the first cycle, discussions with the team of education experts 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 (Section 2.1).

Given the emphasis on providing highly realistic visuals, the 3D computer graphics were also developed during the first year for IVT 1.0: two virtual classrooms and thirty unique virtual students were created. As indicated in Figure 3.2, 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 (six cycles of evaluation and refinements for the classroom and four cycles for the students) (discussed in details in Section 3.7.2). 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 complete 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 Figure 3.2, usability questionnaires, CTA protocol and semi structured interviews were used to collect data from education majors, as described in details in Section 3.7.3.

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 3.1. 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 five 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 New-Jersey, as described in Section 3.7.4. The current IVT-T 4.3 version contains nine vignette scenarios. Refinements of the simulator and of the website were made according the results from the evaluation of IVT-T 3.0. 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 (IVT-T's fourth Research Objective above).

In the next sections, we present our approach to build the IVT-T classroom simulator based on these requirements and on the requirements established from the related research.

3.2 Overview

IVT-T is composed of a virtual classroom simulator application and a website. By connecting on the IVT-T website, ECTs have unlimited access to practice sessions. Practice sessions consist of playing the scenarios 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 on three 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 can practice autonomously. To encode vignettes, we used the MASCARET [Querrec et al., 2004] which associates sequences of virtual students' actions within the 3D environment using UML concepts (Figure 3.6).
- **3D Virtual Students**: By describing realistic scenarios, vignettes describe two types of disruptive students' behaviors (Off-task and aggressive). The

behaviors are impersonated by 3D virtual student. The IVT-T system counts 30 unique students designed to reflect appropriate ages.

- **3D Classrooms**: The virtual students are displayed in two 3D classrooms (1^{st}grade and 6^{th}grade). To ensure ECTs' immersion in the classrooms and foster the learning, efforts were concentrated on different physical arrangements reflecting the academic level, quality, and realism.

The simulator is available online on the IVT-T website to guarantee access from anywhere and it can be run on any computer with a graphic card supporting 3D, thus allowing a broad range of possible users (24/7 usage).

3.3 Architecture

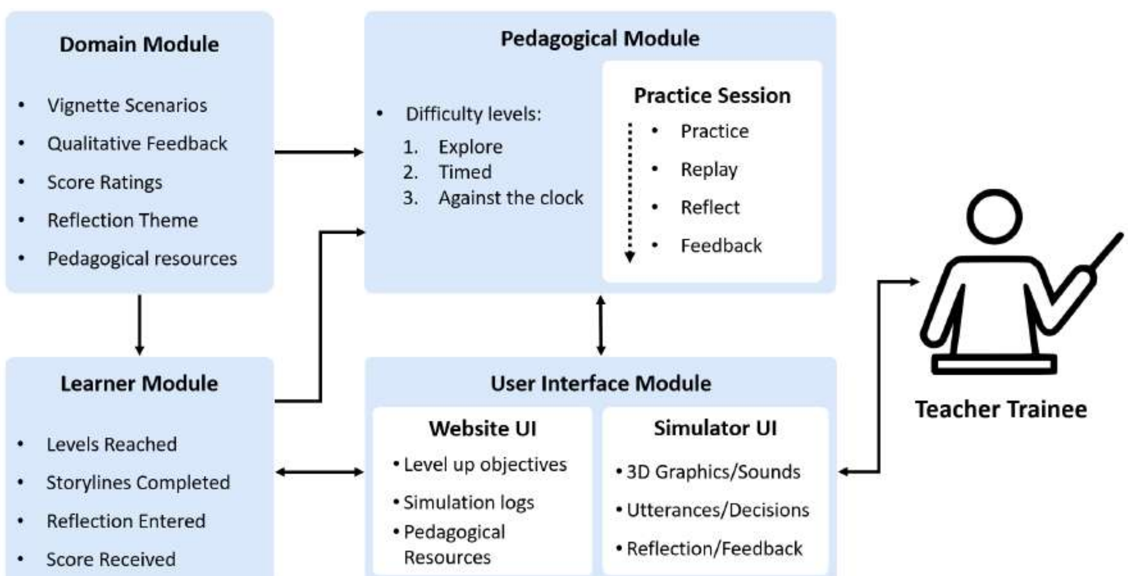


Figure 3.3: IVT-T High-Level System Architecture. IVT-T's architecture is based on Intelligent Tutoring Systems architecture.

The architecture of the IVT-T system contains the four main components of an Intelligent Tutoring System (ITS) (Figure 3.3): (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 (realistic scenarios of teacher-student interactions) with their respective scores and feedback for each decision. The type of ending (positive, mixed or negative), i.e. the quantitative description of situations reached after the sequence of decision made by ECTs is also included in the vignettes' content. The domain module also contains reflection theme used in the IVT-T's practice sessions and pedagogical resources. The domain modules communicates the vignettes' content to the pedagogical module which organize them by difficulty levels and the pedagogical module uses the reflections themes, qualitative feedback and score ratings during the practice sessions. The domain modules also communicates to the learner module the qualitative feedback and the scores of the decision made by the users.

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

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 (level 1), e.g. they can choose to make the worst decision to see what happen. In level 2, the system starts keeping track of their scores, an incremental counter gives them an idea of how long they took before making a decision. Therefore, ECTs can follow their progress as they complete practice sessions. Moreover, they need to 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. The IVT-T system will pick the worst possible decision for them if they have not made a selection.

2. **Replay**: After the practice, ECTs watch the replay of the practice. During this phase they can start evaluating their own performance and reflect on the choices they made.

3. **Reflect**: ECTs are encouraged to reflect on the decisions taken during the practice phase. The domain module includes reflection themes that contain different questions to guide ECTs' reflection. The number of themes and questions can be expanded.

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 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% *have been accomplished"* or *"four different storylines have been accomplished"* to reach more complex simulation with more challenging type and intensities of behaviors. The pedagogical module plays the practice session sequence within the User Interface module.

The Learner Module: To assess user progress and learning, users' current knowledge needs to be represented in the system. In IVT-T, 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. Finally, reflections made during the practice session are aggregated to the Learner module and accessible from the training log. The leaner module communicates the storylines completed to the pedagogical module which, depending on the difficulty level conditions, display the available vignette scenarios to the user through the user interface.

The UI Module: In an ITS, the UI module establishes how information is presented to users. The IVT-T system contains 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. In the website, ECTs can review user storylines, scores, and reflections. Moreover, users also have access to other pedagogical content such as disruptive students' biography or online courses. The UI module, receiving inputs from the teacher trainee, connects to the pedagogical module to display the different phases of the practice sessions. Additionally, the UI module gather the data of the user through the learner module and displays them to the trainee on the website.

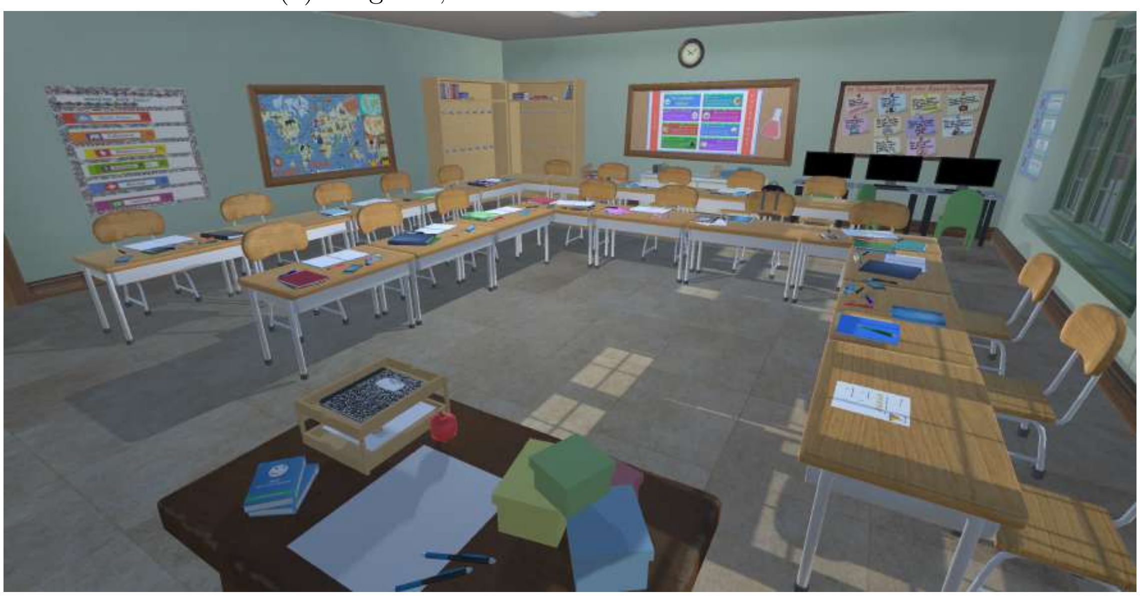
3.4 IVT-T's Graphics

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

3.4.1 3D Classrooms

Two 3D classrooms were designed for IVT-T, one 1^{st} grade and one 6^{th} grade. Significant effort was made to ensure realism of the classrooms to enhance ECTs' immersion, and a number of iterations of prototypes and feedback from education

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



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

Figure 3.4: IVT-T Virtual Classrooms

experts 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 1^{st} grade classroom (Figure 3.4a) for example, a rocking chair and a carpet were added to the corner, and the tables were organized into clusters. On the other hand, in the 6^{th} grade classroom (Figure 3.4b) computers were added, rather than the alphabet, scientific methods are displayed on the walls, and 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 indicates when to play these additional ambient sounds.

3.4.2 3D Virtual Students

Virtual students were developed using MakeHuman [MakeHuman, 2014], 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 clothing, 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 1^{st} grader and 15 6^{th} grader were designed. Each character has a unique skin color, hairstyle and a body shape, illustrated by figure 3.5 showing the face of three 1^{st} grade students.

Figure 3.5: IVT-T Virtual Students version 4 - 1^{st} grade students

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

First, the list of all possible behaviors had to be extracted from the vignettes that were provided by the education experts. For economical reasons and in order to save time, we reduced this number by splitting the animated body parts of the virtual students and by reusing the same animation for different behaviors. For example, behaviors *Point to board* and *Take from teacher sitting* involve the same body movements with different hands disposition. Thus by applying different hands movements with different body posture, we narrowed down the number of needed animations. Freely accessible online databases provided animations exhibiting common behaviors such as *Walking* or *Take from teacher standing*. Finally, behaviors that were too specific to IVT-T such as *Knock desk over* or *Middle finger to the class* were recorded using two Kinects in stereo [Gao et al., 2015] combined with a software linking depth maps with virtual humanoid skeleton (Ipisoft [iPi Soft LLC, 2020]).

Because some behaviors had irregular movements and were colliding with objects in the environments (e.g., chairs and desks), we refined animations by adjusting body positions and by smoothing movements over time to fit the environment and to look more natural.

The list of behaviors recorded is shown in Table 3.1. Some behaviors required to be recorded for both the standing and the sitting position. For example, *Writing* includes writing on the book or writing on the board. Similarly, some behaviors were recorded for a one hand version (taking a pencil from inside the desk) and two hands version (taking a book from inside the desk).

Neutral classroom behaviors			
Idle sitting (4)	Writing*	Point to board	Open/Close book
Idle standing	Reading	Point to paper	Flip pages
Stand up	Rummage desk	Twist on chair	Slide book
Sit down	Rummage backpack*	Take/Put chair	Sharpen pencil
Walking	Take/put on/in desk**	Open/Close door	Work with neighbor
Raise hand		Walk with chair	
Scoot chair			

Off task behaviors		**Aggressive behaviors**	
Low intensity	*High intensity*	*Low intensity*	*High intensity*
Cover mouth	Doodling	Tap pencil	Knock desk
Plop on chair	Draw on hand	Play with phone	Push book
Shrug shoulders**	Rocking chair	Finger tapping	Push chair
Lean back	Roll pencil	Wave hand	Slam book
Slouching	Spill paint		Slam door
Elbow on desk	Play with paint		Drum on desk*
Head on desk	Play with Ipad		Foot kick
Head on arms	Cross arms*		Singing*
Whisper to neighbor	Listening to music		Middle finger
			Drop Ipad
			Take out phone

Table 3.1: List of Virtual Students Behaviors. List of behaviors (neutral classroom behaviors, off-task behaviors, and aggressive behaviors) displayed by the virtual students in IVT-T. ** indicates that the behavior exist for the standing and the sitting position; ** indicate the the behavior exist for 1 hand and for 2 hands.*

3.5 IVT-T Simulator

The main component of the simulator is the graphics, including the classroom 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 ECT users. This section describes the vignette component of IVT-T and how vignettes scenarios are translated into a 3D simulation.

3.5.1 Vignette

A total of nine vignette scenarios were created in LucidChart [LucidChart, 2020] (Left of Figure 3.6). The scenarios are represented as decision trees following a hierarchical structure. 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. IVT-T includes two main types of behaviors:

- Disruptive behavior: A disruptive virtual student can be off-task (OT), daydreaming or looking out the window while all other students are reading, or aggressive/non-compliant (A/NC), refusing to follow instructions or exhibiting aggressive behaviors (verbal and/or physical). From the set of characters generated for each classroom, a subset of four disruptive characters was implemented, OT 1^{st} grade and 6^{th} grade students, and A/NC 1^{st} grade and 6^{th} grade student.
- Non disruptive behavior: Non disruptive agents are controlled by a finite state machine, looping through behaviors relevant to the context of the vignette scenario such as reading, writing, looking at the board, turning pages.

A special type of node, decision nodes (yellow diamond-shaped box in Figure 3.6), describes where ECTs need to make a decision to advance in the scenario. Generally, to reach an end node (rounded red box in Figure 3.6) ECTs go through three to eight decision nodes. A storyline is completed when an end node is reached.

3.5.2 3D Simulation

By going through the vignettes, ECTs face different classroom situations so they can intensively practice their behavior 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.

3.5.3 From vignette to 3D simulation

In order to create a simulation, vignettes are translated into a 3D interactive environment. However, vignettes can be very large: on average, each IVT-T vignette contains around 50 decisions segments like the one illustrated in Figure 3.6, with each pathway containing between 10 to 20 nodes. Each node represents an utterance and/or an action to be performed by one or many virtual students.

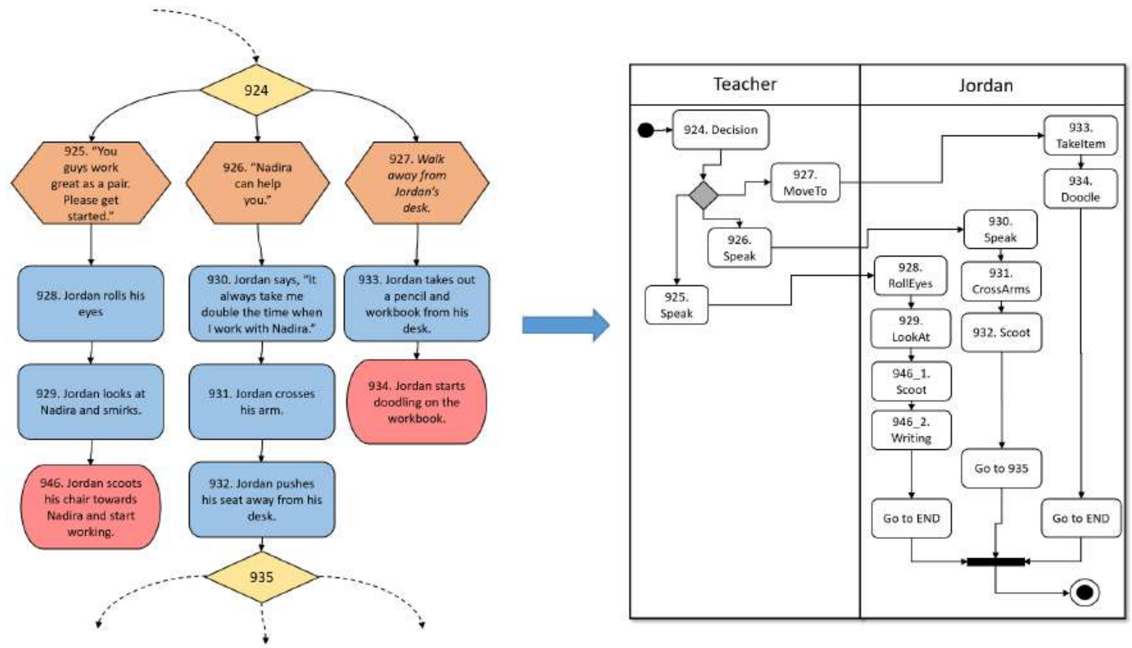


Figure 3.6: Vignette to UML Activity Diagram. Translation of a vignette sample to a UML Activity Diagram. Diamonds are interpreted as decision-merge nodes. Action nodes for a given virtual student are attributed to its corresponding role in the activity diagram.

We decided to use a multi-agent framework called 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 which are graphical representation of a workflow of actions realize by one or many roles. By drawing parallels between these two representations, as vignettes contain sequences of actions with different actors, we were able to integrate these massive scenarios inside a 3D environment using MASCARET.

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 (Figure 3.6).

The method of translating LucidChart directly into a 3D simulation present two 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.

3.6 IVT-T 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 (Figure 3.3), 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.

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

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

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

Figure 3.7: IVT-T 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.

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 (Figure 3.7) is also displayed at the bottom of the screen to maintain consistency with the ECTs' utterance display. Up to three 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.

Additionally, to enhance the ECTs immersion in the environment, actual children were recorded saying the phrases of the utterances. By using the audio source

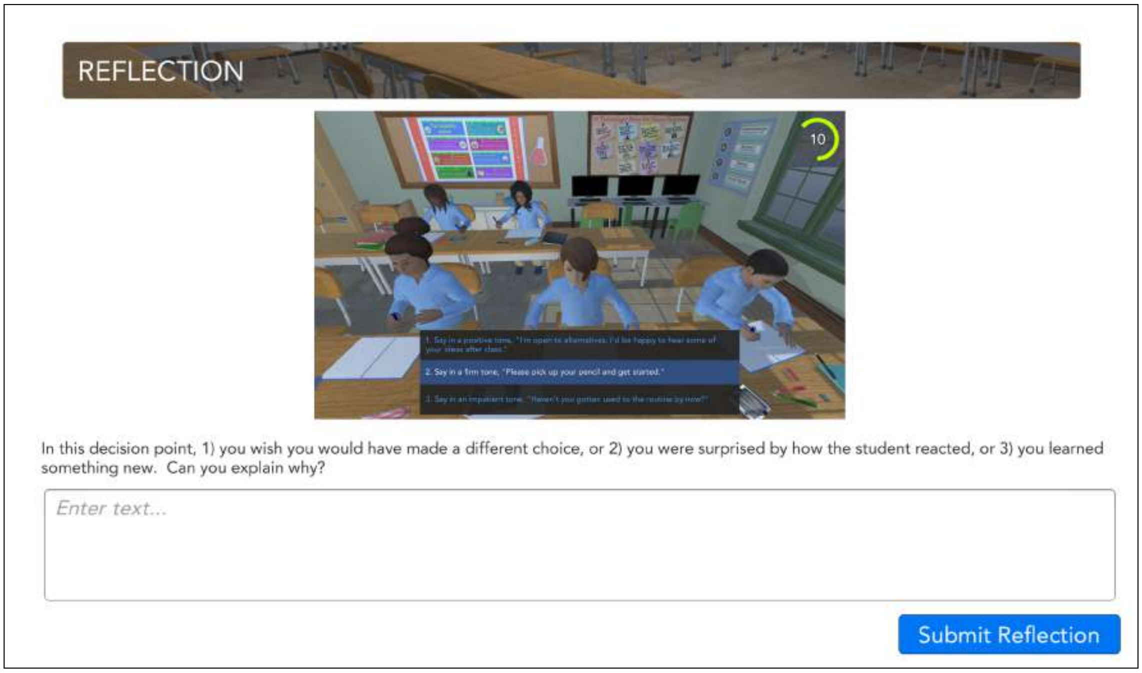


Figure 3.8: IVT-T Reflection. After selecting a decision to reflect on, teachers are invited to answer a question to guide their reflection. A screenshot of the decision is displayed to help teachers consider the choices they had.

component in the Unity 3D engine, 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.

During an IVT-T practice session, ECTs also reflect on their decisions. During the reflection phase, ECTs select a decision to reflect on and are then prompted with a question to guide their reflection (Figure 3.8).

Finally, 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 (Figure 3.9). 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.

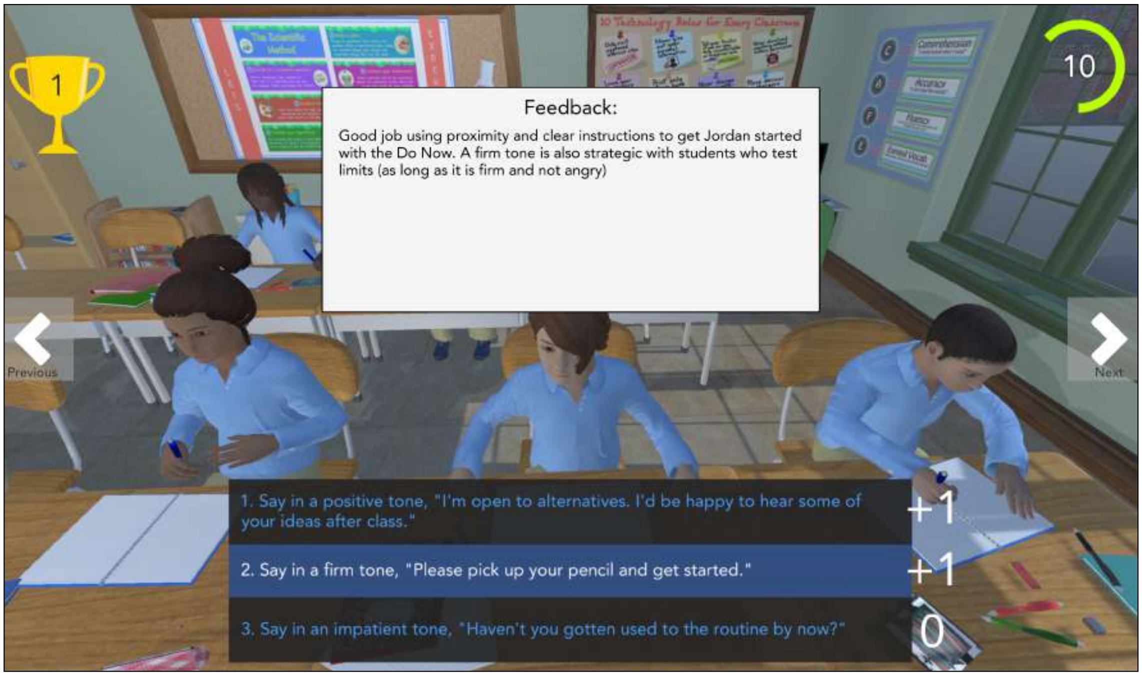


Figure 3.9: IVT-T Feedback. After the reflection phase, ECTs visualize in sequence screenshots of their decisions with corresponding quantitative and qualitative feedback. To move through the sequence of decision screenshots, ECTs use the arrows on the side of the screen.

On the IVT-T website, ECTs can download the simulator application to practice directly from their computer. Moreover, ECTs can review all the practice sessions they completed. For each session, they can visualize a storyline summary showing the number of effective decisions made and the strategies used (Figure 3.10). For each decision, they also have access to the choice they made, the feedback and score they received and the reflection they entered. Additionally, ECTs can review the objectives they need to achieve to gain access to more complex levels.

The website also gives access to biographies of the disruptive virtual students (e.g. *Who are they? What are their relationships with classmates and family?*) and other pedagogical contents are available in the domain module.

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*

Figure 3.10: IVT-T 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.

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 (Figure 3.11).

3.7 IVT-T Requirements' Evaluation

This section describes the evaluation of vignette scenarios to determine their authenticity and realism as well as the evaluation of the 3D classrooms and virtual students. The second part presents the usability study of IVT-T. The results in this section were presented in published work [Shernoff et al., 2018, Shernoff et al., 2020].

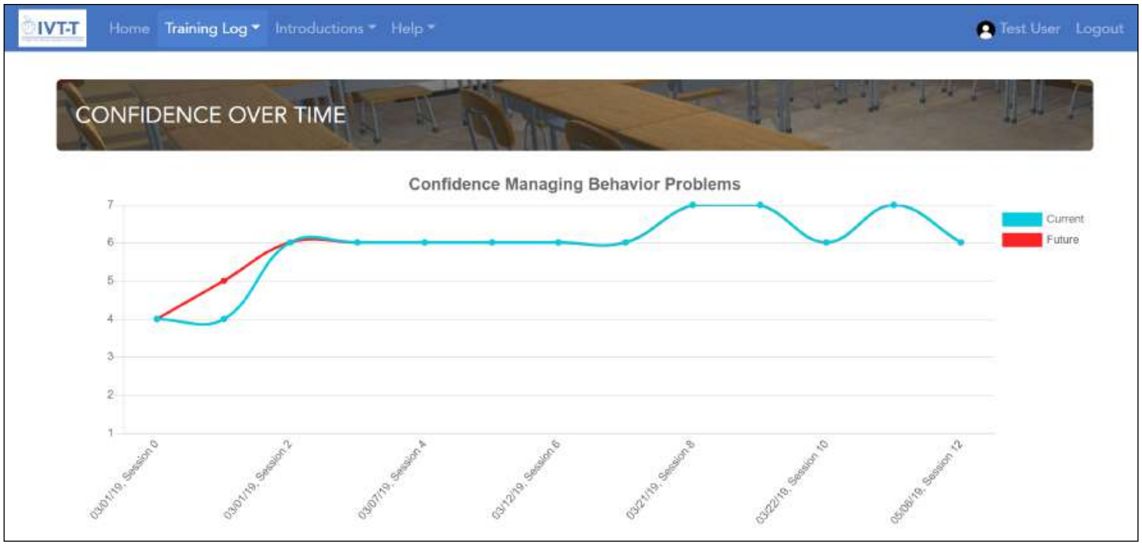


Figure 3.11: IVT-T Confidence Ratings. TSQ answers [Webster-Stratton, 2005]. The graph shows the evolution of ECTs' confidence in managing disruptive behaviors in their classroom currently and in the future.

3.7.1 Behavioral Fidelity: Vignette scenarios

To deliver realistic and engaging content to ECTs, IVT-T scenarios were assessed in terms of realism, consistency, and engagement. A total of twelve vignettes were evaluated by the advisory board: three scenario levels for four virtual disruptive students. Only nine vignettes (three students) are included in the final implementation of IVT-T.

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.

Procedure

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

Measures

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 (1^{st} grade and 6^{th} grade). The advisory board also evaluated how engaging the vignette scenarios were using the same 4-point scale. Qualitative feedback consisted of changes proposition to improve the scenarios.

Apparatus

A total of five vignette prototypes, for each individual 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 Figure 3.6).

Vignette scenarios evaluation results

Results show that the earlier versions of the vignettes were rated lower than revised versions for each evaluation category (Figure 3.12). Vignettes depicting aggressive behaviors (Mean = 3.76, SD = .36) were generally rated more engaging than off-task behaviors (Mean = 3.44, SD = .39). Overall realism and logic ratings of the scenario by the experienced teachers are high, 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 really liked the paths in general and I thought that the students' reactions and behaviors were pretty typical for first graders. However, I thought some of the teacher options were too harsh for a first grade teacher. I also thought that some decision point options inconsistent."*. This type of feedback helped direct the refinements in the vignettes and generate more realistic situations.

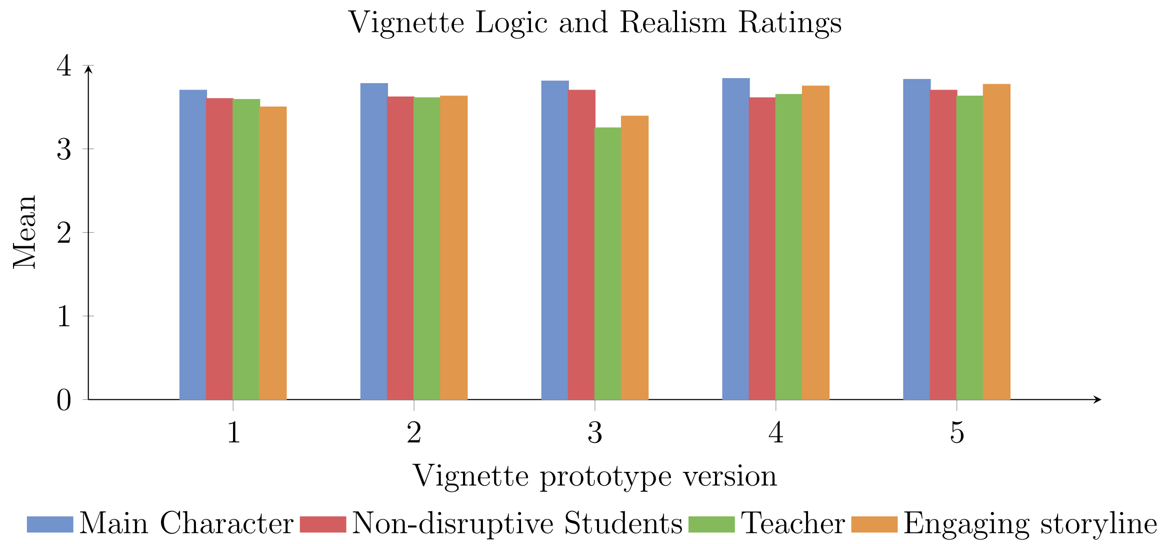


Figure 3.12: IVT-T Vignettes Logic and Realism. Ratings for the 5 evaluations (1 = Strongly Disagree - 4 = Strongly Agree). Vignettes were refined between each evaluation.

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

3.7.2 Environment Fidelity: 3D Classrooms and 3D Virtual Students

To address the environment fidelity requirement, two virtual classrooms and 30 virtual students were evaluated in terms of authenticity, realism and representativeness.

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.

Procedure

A total of six classrooms prototypes (for both the 1^{st} grade and 6^{th} 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 four prototypes were proposed for each of the 30 students.

Measures

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, student's 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 = Fair, 3 = Good and 4 = Outstanding).

Qualitative feedback consisted of recommendations on physical characteristics for each prototype.

Apparatus

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

3D Classrooms evaluation results

Quantitative ratings are shown in Figure 3.13. 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 student evaluation results

Figure 3.14 shows the mean of the 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 "outstanding", only two characters were rated as "fair".

The advisory board provided qualitative feedback to express what they liked 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 graphic realism for VTEs could be found, it is not possible to compare the realism of IVT-T's graphic to other systems. As the impact of graphics quality on the learning outcomes of VTEs is still ongoing research, we encourage existing and future VTEs to proceed to similar graphic evaluation in order to provide comparative values to better study the effect of graphics quality for VTEs.

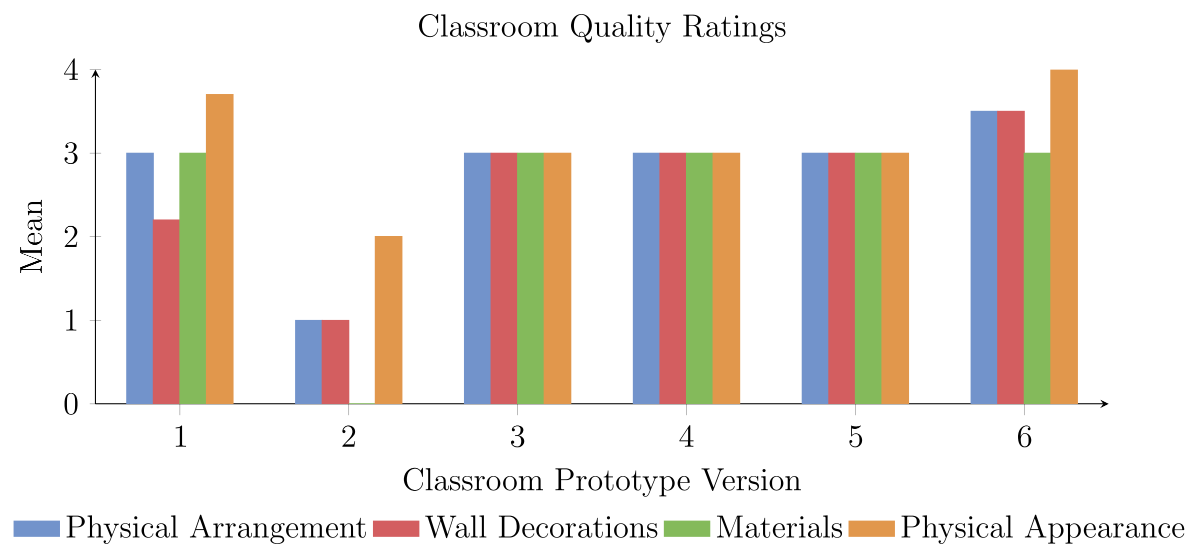


Figure 3.13: IVT-T Virtual Classrooms Realism. Quality rating means of the 2 virtual classrooms for the 6 prototypes (1 = Poor - 4 = Outstanding). Virtual classrooms were refined between each evaluation.

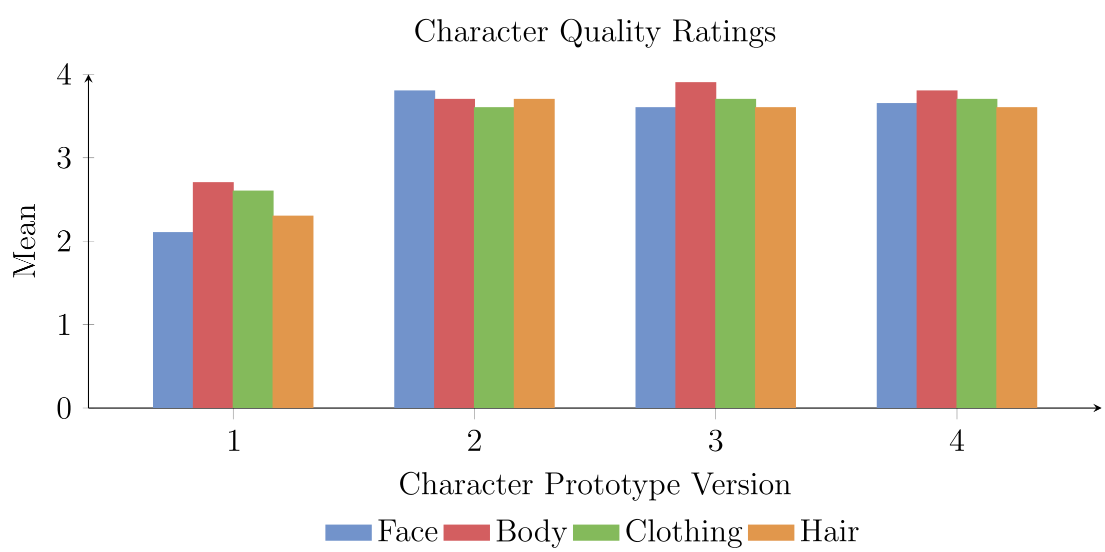


Figure 3.14: IVT-T Virtual Students Realism. Quality rating means of the 30 virtual students for the 4 prototypes (1 = Poor - 4 = Outstanding). Virtual students were refined between each evaluation.

3.7.3 Interactivity: Usability and UX

We designed a study to assess IVT-T usability and 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 use of the system. The evaluation focuses on IVT-T's learnability and efficiency.

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.

Procedure

Each 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 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 [IJsselsteijn 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).

Measures

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 explores whether teachers' gaming experience influences use of technology by asking questions about gaming habits and gaming experiences. 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.

Apparatus

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.

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 considered themselves as gamers, the other five play video games for less than an hour a week. Based on the result of this survey, we identified that most of the participants were not gamers and would be able to provide interesting feedback on the usability of the IVT-T system which is meant for use with population not necessarily technologically savvy.

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 (Figure 3.15). 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 SUS questionnaire confirmed the results from the QUIS, as users 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 before they could get going with the system (M=1.57,SD=0.78) neither did 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).

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

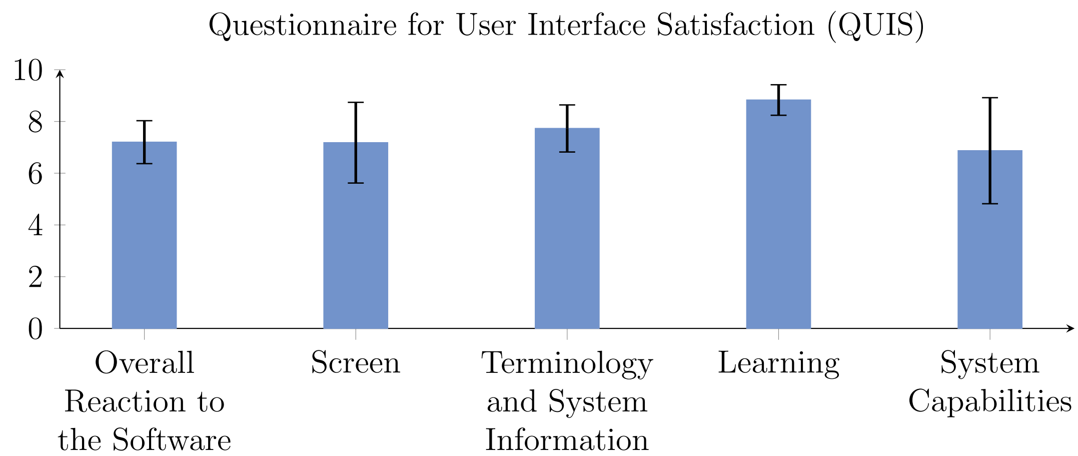


Figure 3.15: IVT-T Usability: QUIS Results. QUIS is comprised of six subscales. Each of these scales includes between four and six items, rated on a 10-point scale (0 = hard to 9 = easy, 0 = confusing to 9 = very clear, 0 = rigid to 9 = flexible). Items in each subscale are averaged to compute a mean score for each subscale.

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."*.

In addition to confirming the usability of the system, the semi-structured interviews supported the realism of disruptive students' behavior and of the audio *"Differently depending on the child but it was pretty accurate, you can tell in his voice that he was pretty sassy and that's good because they are sassy."* or *"he seemed really realistic"*.

Finally, participants were able to make suggestions on how to improve the system such as integrating facial expressions (for the virtual students but also for user to have the possibility to see their own facial expressions) or integrating virtual colleagues that would make the choices users would not want to make.

3.7.4 Instructional Design: Current Findings

The objective of this preliminary study was to evaluate the instructional design and the usage of IVT-T with practicing K-12 teachers (Evaluation of IVT-T 3.0 in Figure

3.2). Only the preliminary results of the instructional design are presented here, IVT-T usage is left for future work.

Population

A sample of practicing K-12 teachers (n=26) were recruited from elementary schools.

Procedure

Participants were given access to IVT-T for a 14 week 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.

Measure

The system recorded logs of every IVT-T training session (Scores, reflection entered, and feedback received). An IVT-T training session goes through the following sequence: (1) Practice; (2) Replay (Optional); (3) Reflection; and (4) Feedback (not available at level 1). User can decide to exit the session at any time, however a training session is recorded if at least the Practice phase is completed, i.e. the teacher reached the end of the scenario. An IVT-T training session is considered complete if the feedback for all the decision have been received by the user.

Apparatus

Participants used IVT-T on their personal computers.

Results

During the 14 week 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 completed with teachers going through all the phases of the practice session (Figure 3.16). Since level 1 do not include a feedback phase, the only phases that can be completed at level 1 are practice and reflect. This result indicates that teachers value the reflection and the feedback phase of the instructional design of IVT-T. By themselves, teachers went through the overall sequence of the IVT-T training session. Additionally, at the end of the practice phase, teachers were given the choice to watch the replay of their actions or to skip directly to the reflection phase. This preliminary study also showed that teachers never watched a replay of their simulations.

The analysis of the reflection phase shows that, most of the time, teachers wrote down their thoughts about the simulations: only 81 reflections (8%) were left blank. A theme that emerged from the reflection was the connection between IVT-T and real classroom experience, 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 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 preliminary 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.

3.8 Discussion

We answered our first research question (**RQ1**) by directing the development of IVT-T with 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

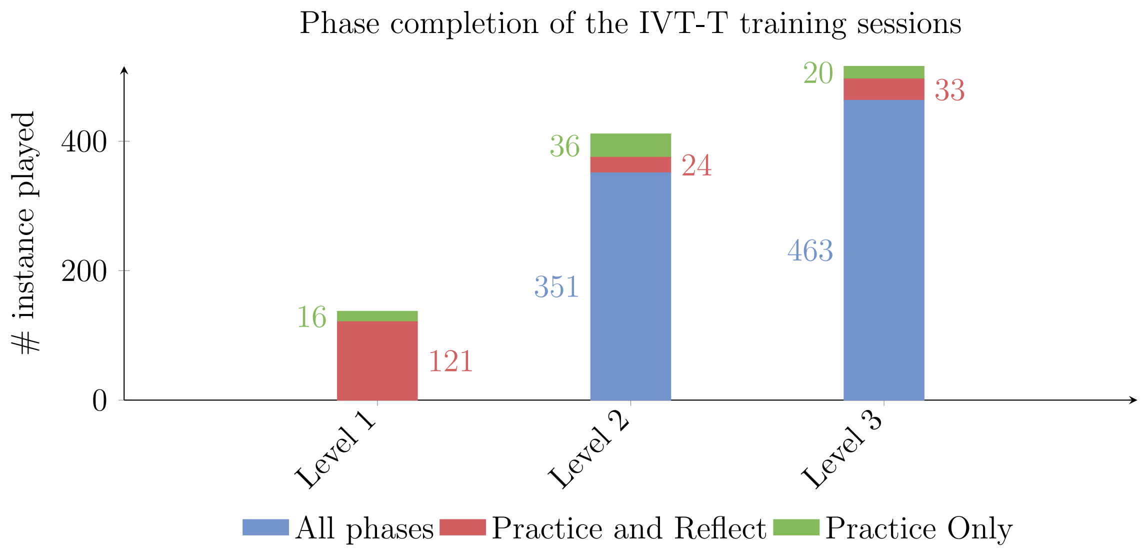


Figure 3.16: IVT-T Practice Sessions Completion. Completion of the different phases of IVT-T training sessions. All phases include (Practice, Reflect and Feedback). Level 1 vignettes do not include the feedback phase.

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. IVT-T integrates 30 virtual students impersonating 15 1st graders and 15 6th graders presenting a realistic number of students per class, compared to the TeachLive simulator [Dieker et al., 2015] which presents a classroom of six 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 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 into the instruction increases the transfer of learning. Finally, ECTs receive automatically generated explicit feedback on the choices they made which resulted 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, thus facilitating ECTs to practice 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 [Dalgarno and Lee, 2009, Bossard et al., 2008] which will result in more transfer of learning. Whereas others argue that, too much fidelity would fall in the uncanny valley and would 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 student 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 their 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.

CHAPTER 4

CONTRIBUTION: METIS - ADVANCED VR AND AFFECT VERSION

METIS is built on the same architecture and uses the same components as IVT-T. However modifications were made to port METIS to immersive VR technologies and to support the display of emotions by METIS virtual students (Figure 4.1). First, a device manager was added to adapt the UI and the interactions to the device used. Secondly, the simulator now integrates a Behavior Markup Language (BML) realizer to allow virtual students to use the BML to display emotions.

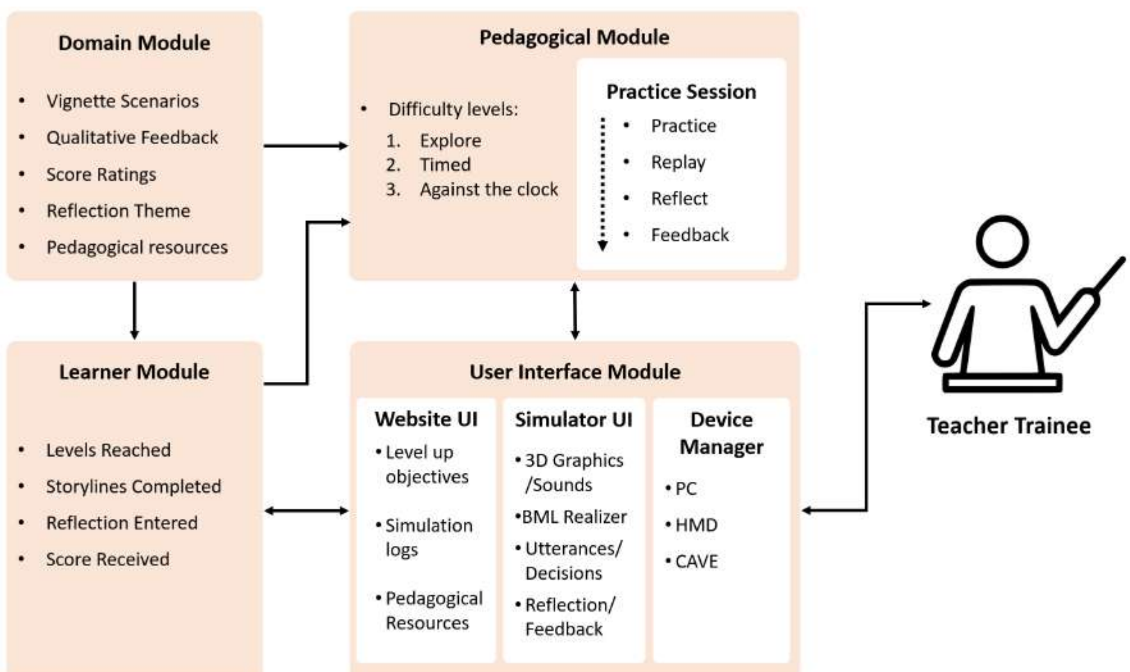


Figure 4.1: METIS High-Level System Architecture.

We present the modifications made to the UI in the first part of this chapter. Secondly, we describe how we enabled the virtual students to display affective non-verbal behaviors using BML. Then, we present the results of our Usability and UX evaluation of METIS for all VR platforms. Finally, we detail how we evaluated the effects of VR platforms and emotional virtual students on users.

4.1 METIS' VR Platforms User Interface

The quality of its UI is one of the main factors for users' acceptance of a system. A challenge with the METIS UI is that it should be equally usable across the three different VR platforms (PC, HMD, CAVE) we considered. With immersive technology however, traditional on-screen graphical UI control elements cannot be used. Therefore, we designed a single METIS UI compatible with desktop/laptop setup and immersive VR platforms.

The METIS simulator enacts the scenarios within the 3D classroom through virtual students' spoken utterances and actions, and it additionally needs to display the decisions to allow for user input. We identified three main types of interactions:

1. Reading virtual students' utterances
2. Making a decision in the scenario to progress to subsequent student behaviors
3. Reading out loud the teacher's utterances

Virtual student utterances. To display utterances of the virtual students and of the teacher (interactions 1 and 3 above), we resorted to 3D speech bubbles located near the speaking character (Figure 4.2) (e.g. above the head for students' utterances and in front of the camera and slightly below the center of the field of view for teacher's utterances). An issue that appeared was the size of the speech bubble being too big for the computer screen when the student was close to the teacher. This problem was addressed by reducing the size of the bubble when the student-teacher distance would drop under a fixed threshold. A word-per-minute ratio was used to

Figure 4.2: METIS UI: Student Speech Bubble. Speech bubble displayed by the student interacted with.

measure the duration of the display of each utterance to ensure a flowing scenario and to leave enough time for users to read and hear the utterances. Finally, to indicate the state of the system to the users, a loading bar was added at the top of the speech bubble to dynamically display the remaining reading time before the next system action.

User's decision. The main interaction with the METIS system is for users to address the disruptive behaviors by selecting one of three options (interaction 2 above) (Figure 4.3). Options are displayed horizontally so as to not block the view of the classroom, and the users can observe the disruptive student to inform their decision. A bright blue laser pointer was used for selection (shown in Figure 4.3).

Teacher trainee's utterances. In order to practice speaking in front of a classroom, METIS users are encouraged to read the utterances out loud.

Navigation. The navigation and orientation of the students and of the teacher are controlled by the system in order to always have the camera facing the action. When an option presented to the user involves movement (*"Go to Jordan's desk and say..."*), the simulator automatically steers the camera to the indicated location in the classroom.

By creating a UI compatible with all three platforms (PC, HMD, CAVE), the device manager only needs to process the user inputs from the controllers specific to each platform.

4.1.1 Differences between VR platforms

There are three main differences between the VR platforms (PC, HMD shown in Figure 4.4b, and CAVE shown in Figure 4.4a): (1) the interaction controller; (2) the navigation technique; and (3) the camera orientation.

Figure 4.3: METIS UI: Decision. As users can still observe the students, they can consider the different options and highlight them with the bright blue laser pointer.

The differences in interaction controllers are due to technological constraints. The CAVE uses an Xbox controller, the HMD uses the HTC Vive controllers provided with the HTC Vive headset, and finally the PC version of METIS uses a regular mouse.

The second difference is the navigation technique. METIS uses teleportation (instantaneous transportation to the target location) for the HMD platform. Steering (smooth transition from the current position to a target location) causes more cybersickness than teleportation in HMDs [Christou and Aristidou, 2017]. Steering however was used for both the CAVE and the PC platforms. Steering can also be responsible for cybersickness in a CAVE-like environment, however it is attributed to the amount of total movement [Ragan et al., 2012]. In METIS, movements are limited to few occasions, therefore steering was also selected for the CAVE.

Finally, the different nature of each VR platform forces the METIS simulator to manage the orientation of the camera in different ways. For the PC platform, the orientation of the camera always faces the action, i.e. users do not control the camera, the system does. This ensured that users would not miss any actions performed by the students. For the HMD and the CAVE platforms, the orientation of the camera cannot be controlled by the system because users can decide to rotate their head

(a) METIS CAVE

(b) METIS HMD

Figure 4.4: METIS CAVE and HMD

to look around. Therefore, for these platforms, only the position of the player is controlled by the system.

4.1.2 Tutorial

As suggested by Zaidi et al. [Zaidi et al., 2019], a tutorial level was created for users to learn independently how to interact with METIS and to avoid any potential introduction of biases by experimenters explaining how to use the system. The tutorial takes place in an empty classroom and consists of three tasks: (1) The first task is to take control of the laser pointer. A text informs users to point the laser to a panel (Figure 4.5). (2) The second task teaches how to select the panel. (3) At the beginning of the third task, a bell rings and students appear at their desks. An introduction text explains the context of the classroom and gives information on the disruptive student's behavior. The interaction with the disruptive student starts when participants click on the panel *"Start"* below the introduction text.

Figure 4.5: METIS Tutorial. Explaining users how to interact with button panels.

4.2 METIS Virtual students display of emotions

In METIS, the disruptive virtual students can display facial expressions. A set of seven facial expressions (Happiness, Sadness, Fear, Anger, Embarrassment, Contempt, Boredom) were selected according to the situation described in the scenarios (Figure 4.6). As the action in the scenarios specifies which 3D animations should be played, facial expressions are also indicated when they need to be played in the scenarios.

4.2.1 Facial expressions animations

Facial expressions were developed using MakeHuman [MakeHuman, 2014] facial rig. The bones of the face were adjusted to match the Action Units (AUs) of the different emotions as specified in different studies [Amini and Lisetti, 2013, Scherer et al., 2019]. Action units represent the smallest group of muscles that can move independently in the human face [Ekman, 1997], and they can be simulated with facial animations activating various AUs. We recorded animation of facial expressions by transitioning from a neutral face pose (no AUs activated) to a facial expression pose (the corresponding set of AUs to the facial expression are activated). A total of seven facial expressions animations were recorded, the facial expression poses are shown in Figure 4.6.

4.2.2 Behavior Markup Language

The BML [Kopp et al., 2006, Vilhjálmsson et al., 2007] is an XML description language for controlling the verbal and nonverbal behavior of virtual agents. BML is being used in a variety of application such as job interview practice for young adults with autism spectrum disorder [Hartholt et al., 2019], training medical experts to break bad news to patients [Ochs et al., 2018], and creating online conversational agents for older people [Llorach et al., 2019].

The benefit of BML is that the BML language is independent of any platform. Any system integrating a BML realizer (i.e. the module interpreting BML and generating the specified behavior on a virtual agent) is able to play any behaviors described as a BML. Therefore it allows for a generic way to play verbal and non-verbal behaviors.

BML is used to describe a variety of verbal and nonverbal behavior through its core language using xml tags (`<speech>`: to specify agents' utterances, `<head>`: to specify head animations such as nod, `<gaze>`: to indicate a target object or person to gaze at,`<body>`: to play a body animation,`<gesture>`: to describe arms gesture, `<faceLexeme>`: to play facial expressions animations). BML is also able to synchronizes the animations with the speech by introducing marks in the text, marks

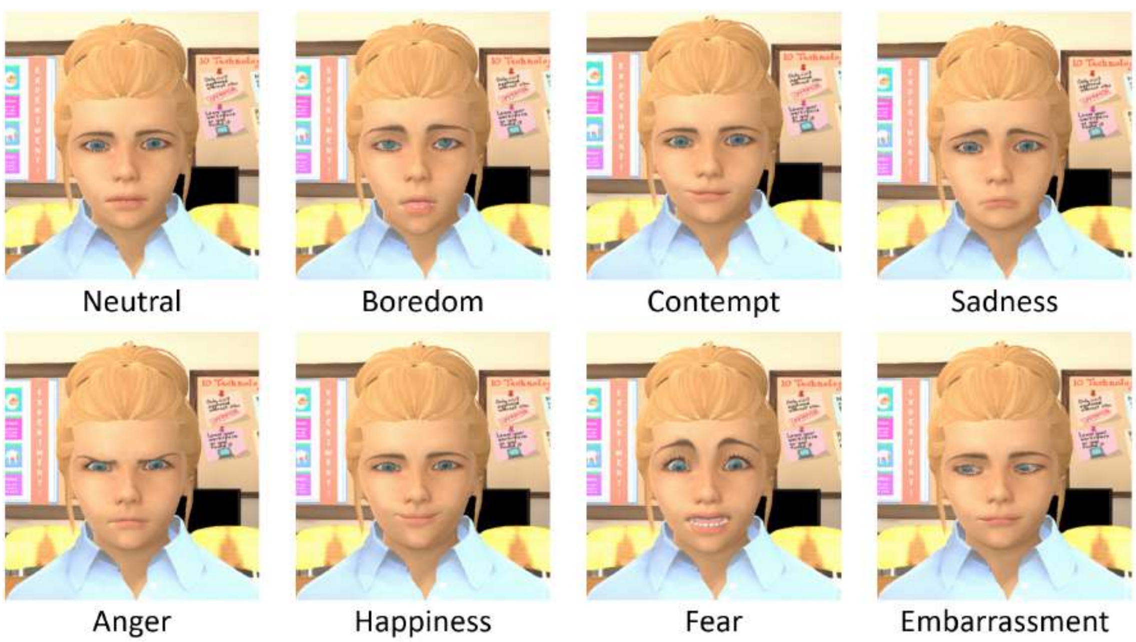


Figure 4.6: METIS Facial Expressions. Disruptive students facial expressions, action units (AUs) are given for each facial expressions [Amini and Lisetti, 2013, Scherer et al., 2019]. Boredom 1+2+25+26; Contempt:12+14RE; Sadness:1+4+15; Anger:4+5+7+23; Happiness:6+12; Fear:1+2+4+5+7+20+26; Embarrassment:12+54+62+64;

indicating key times in the speech that can be use to trigger specific animations at the specified key times [Kopp et al., 2006, Vilhjálmsson et al., 2007].

In METIS, vignette scenarios developed by educational expert and validated by experienced teachers, already specify students' utterances, body postures, and students' gaze. Therefore BML was used to control only the facial expressions of the students and the hands animation (open or fist). To interpret the BML, the METIS BML realizer only consider `<faceLexeme>` xml tags and display the corresponding specified lexeme. For instance `<faceLexeme id="f1" lexeme="Happiness" amount="1.0" start="0" end="7">` will start displaying the happiness facial expression one second after the call to display the facial expressions and will return to a neutral step after seven seconds. The facial expressions will be played at full intensity (METIS only includes full intensity for facial expressions).

4.3 Evaluation of METIS: Usability and UX

In this section, we address **RQ2** by detailing the design of our evaluation of the METIS cross-platform system. The goal of this evaluation is to determine if education students can learn to operate METIS on all platforms and if they experience any difficulties and to collect feedback to identify potential usability issues. We are also investigating if education students would accept such technology for their training and how METIS affect their UX including (cybersickness, judgment and emotions). The results presented in this section are part of published work [Delamarre et al., 2020b].

4.3.1 Methodology

To evaluate the usability of the METIS classroom simulator over the different VR platforms (PC, HMD, CAVE) we conducted a study with 18 education students. We used a between-subject study design to evaluate the usability between different VR platform, similarly to other studies [Cao et al., 2019, McMahan et al., 2012, Sharples et al., 2008]. Therefore, each of our participants were subjected to one VR platform. Our goal was to determine the usability of the system, technology adoption, experience consequence, users' emotion and attractiveness.

Population

Eighteen participants (10 Female, 7 Male, and 1 Other) aged between 18-44 (eleven 18-24 years old, five 25-34 years old, and two 35-44 years old) took part in the study. All participants were students in an education program. On average, each group spent between 1 to 3 hours playing video game and they all reported they felt comfortable using basic computer applications.

Procedure

Once recruited, participants came to the room where they signed an informed consent form, and completed a survey about gaming experience and general demographics information. Then the instructor, following a script, explained the system and the goal of the study. In addition, participants were instructed to verbally express their thoughts and concerns as they interacted with the classroom simulator.

Then, participants interacted with their randomly assigned VR platform. For the CAVE condition, the instructor accompanied the subject to the CAVE room across the hall. For the PC and the HMD conditions, participants remained in the same room. The instructor provided the interaction device and indicated the interaction button (the left click of the mouse for the PC condition, the trigger of HTC Vive controller for the HMD condition, and the A button on a Xbox controller for the CAVE condition). After that, the instructor helped the participant adjust the device (VR headset for the HMD condition, and stereoscopic glasses for the CAVE condition).

Participants interacted with the three virtual students (Jordan, Michael, and Ava) in three scenarios provided in a random order. Before each scenario, they completed the same tutorial described in the previous section.

After participants completed the set of three scenarios, they sat down at a desk in the original room with the instructor to participate in a semi-structured interview. After the interview, participants completed the remaining surveys: QUIS [Chin et al., 1988], Technology adoption, Users' Emotions, Experience consequence, Judgement, and open-ended questions [Tcha-Tokey et al., 2016]. Overall, each session lasted approximately 60 minutes.

Measures

Concurrent Think Aloud (CTA) As participants interacted with the tutorial and the scenarios, they were asked to "think aloud". [Cooke, 2010]. The tutorial gave tasks to help users learn how to interact with system. The tasks consisted of using the laser pointer to hover over panels in the virtual environment, clicking on a demo panel, clicking on the start panel, and making a decision when options appeared. The instructor gave two additional tasks: identifying the disruptive student after reading the scenario introduction, and identify the emotions expressed by the disruptive student. If and when participants stopped talking, the instructor reminded them to keep talking.

The instructor took field notes about usability issues detected by the participants, as well as about elements that triggered positive or negative reaction from the participants.

Semi Structured Interview After the interaction with the three scenarios, subjects participated in semi-structured interviews. Interviews started with general questions such as *What was it like to use the classroom simulator?* and *What did you like the most/least about the classroom simulator? Why?*. Other questions focused on the user interface, the graphics and their perception of the emotions of the virtual students. Finally, questions about the usefulness of the classroom simulator and technology were asked before the interview concluded.

QUIS Questionnaire for User Satisfaction (QUIS) [Chin et al., 1988] is a standardized usability survey measuring the quality and the satisfaction of interactive software. It is composed of an overall index of usability and four subscales (Screen Design and Layout, Terminology and Systems Information, Learning, and System Capabilities). Each Item is rated on a 10-point scale (i.e., 0 = hard to 9 = easy, 0 = inconsistent to 9 = consistent, 0 = confusing to 9 = very clear, 0 = rigid to 9 = flexible). The QUIS has an overall $\alpha = .94$ with average usability ratings from prior research ranging from 4.72 to 7.02 [Chin et al., 1988].

Immersive Virtual Environment Questionnaire - IVEQ The Immersive Virtual Environment Questionnaire (IVEQ) was developed by Tcha-Tokey et al. [Tcha-Tokey et al., 2016]. This questionnaire aims at evaluating the overall UX of immersive virtual environments. We removed components of the questionnaire related to presence, immersion and engagement which were most likely to be affected by the structure of the CTA requesting participants to vocalize their experience (rather than immersing themselves fully). We describe below the subscales that we used from the IVEQ.

- **Technology adoption**: This subscale is composed of six items and was adapted from the Unified Technology Acceptance and Use of Technology questionnaire created by Venkatesh and al. [Venkatesh et al., 2003]. This questionnaire measure the degree to which users will adopt the system and more globally, the likelihood of the system successfully being introduced to end-users.

- **Experience Consequence**: The experience consequence subscale measures the negative effect the system can have on the users such as stress, dizziness, and cyber-sickness. This subscale was adapted from the Simulator Sickness Questionnaire created by Kennedy et al. [Kennedy et al., 1993] and contains eight items.

- **Users' Emotions**: The user emotion subscale measures the user's self-reported feelings in the virtual environment (e.g. joy, pleasure, satisfaction, frustration). It was adapted from the Achievement Emotions Questionnaire [Pekrun et al., 2011]. Items were selected to fill relevant emotions categories; three emotions were selected (positive activating: enjoyment; negative activating: anxiety; and negative deactivating: boredom).

- **Judgement**: Judgement describes the system's attractiveness to users. It is based on the AttracDiff questionnaire created by Hassenzahl et al. [Hassenzahl et al., 2003]. Subscales of the AttracDiff questionnaire concerns perceived pragmatic quality, perceived hedonic quality-stimulation, perceived hedonic quality identification, attractiveness. This subscale was assessed using five items.

Apparatus

For the PC and HMD condition, the classroom simulator was launched on a Corsair 64bit with 16GB of RAM, a Intel processor CPU i7-4790k 4.00GHz and a NVIDIA GeForce GTX 1080 Ti graphics card. The operating system was Windows 10 Education. The sound was played from a Dell A525 2.1 speaker system. The HMD used is a HTC Vive with its respective controllers. The CAVE is a 5 wall hexagon shaped (5.7m $\times$ 6m). It includes real-time head and controller tracking and active stereoscopic display (Figure 4.4a).

4.3.2 Results

In this section, we present the usability results obtained for each VR platform and the usability issues observed during the CTA. Then, we present the results of the technology adoption, experience consequence, users' emotions, and judgment questionnaires.

Usability: Quantitative

The Questionnaire for User Satisfaction [Chin et al., 1988] allows a comparison across the three platforms (average QUIS ratings from previous study range from 7.08 to 7.94; [Su et al., 2019]). The ratings for the QUIS questionnaire for each platform is presented (Figure 4.7). The overall reaction to the METIS classroom simulator rated higher than average (CAVE: $M = 8.99$, $SD = 0.82$; PC: $M = 7.64$, $SD = 1.71$; HMD: $M = 9.33$, $SD = 0.69$). The system learnability was also rated as high for the three conditions (CAVE: $M = 9.00$, $SD = 1.2$; PC: $M = 8.5$, $SD = 1.28$; HMD: $M = 9.33$, $SD = 0.81$). However, an item of the Screen subscale about the ease of reading characters on screen revealed that PC users experienced difficulties with reading texts on the screen (PC: $M = 6.50$, $SD = 2.67$; CAVE: $M = 9.67$, $SD = 0.52$; HMD: $M = 8.00$, $SD = 1.67$).

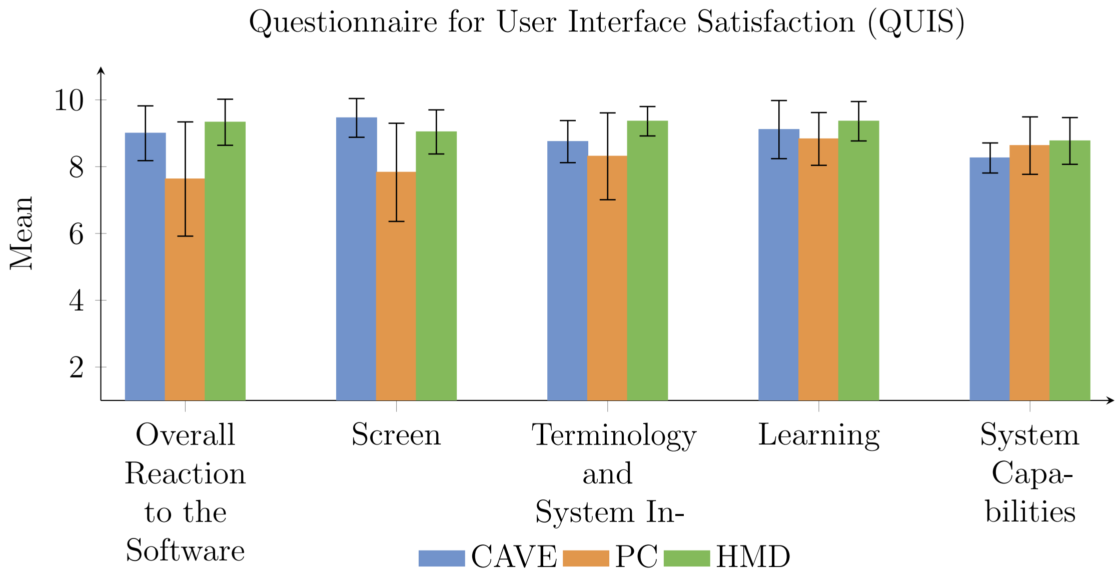


Figure 4.7: METIS Usability: QUIS Results. **QUIS** is comprised of six subscales. Each of these scales includes between four and six items, rated on a 10-point scale (0 = hard to 9 = easy, 0 = confusing to 9 = very clear, 0 = rigid to 9 = flexible). Items in each subscale are averaged to compute a mean score for each subscale.

Usability: Qualitative

The CTA helped detect usability issues that were present in all platforms. However, no usability issues prevented users from continuing their interactions with the classroom simulator.

At the beginning of each scenario, a paragraph introduced the classroom situation to the participants for each disruptive student. After reading the introduction and before they could go on with the simulation, the instructor asked the subjects to point to the disruptive students with the laser pointer. Whereas participants had no difficulties identifying Jordan and Michael (respectively 100% and 89% of correct identification), only one was able to correctly identify Ava (5.6% of correct identification). However, after the first utterance pronounced and displayed for Ava, participants realized which student was causing the disruption.

Additionally, 33% of participants did not realize they had to select an option when the scenario reached a decision point. After a moment, by moving the laser pointer they realized they could interact with the option panels and were able to select an option. The problem was not with the action of selecting an option, rather with the understanding that the system was waiting for user input.

At some point during the simulation, and depending on the choices selected by the participants, the virtual student moves to the back of the classroom. In these cases, participants indicated, 100% of the time, that the text on the speech bubble was too small to be read. Moreover, two participants (11%) indicated that sometimes, teacher utterances would disappear before they had time to read it.

The CTA also revealed issues specific to each VR platform. For instance, after completing the tutorial without any problem, two PC participants faced difficulties pointing the laser pointer to the start button below the introduction. After a few tries moving the mouse widely, they managed to continue with the simulation.

For the CAVE condition, two participants expressed that they felt dizzy when the simulator rotated the view to face the disruptive student. They both indicated that the rotation was too fast.

UX: Technology adoption

Figure 4.8 presents the results obtained for the Technology adoption questionnaire. Items are presented as statement and participants indicate how much they agree to each item (Strongly disagree = 1; Strongly agree = 10). Statements are detailed in Figure 4.8. Overall, participants for the three conditions agreed with the statement *"Learning to operate the virtual environment would be easy for me"* (CAVE: $M = 9.50$,

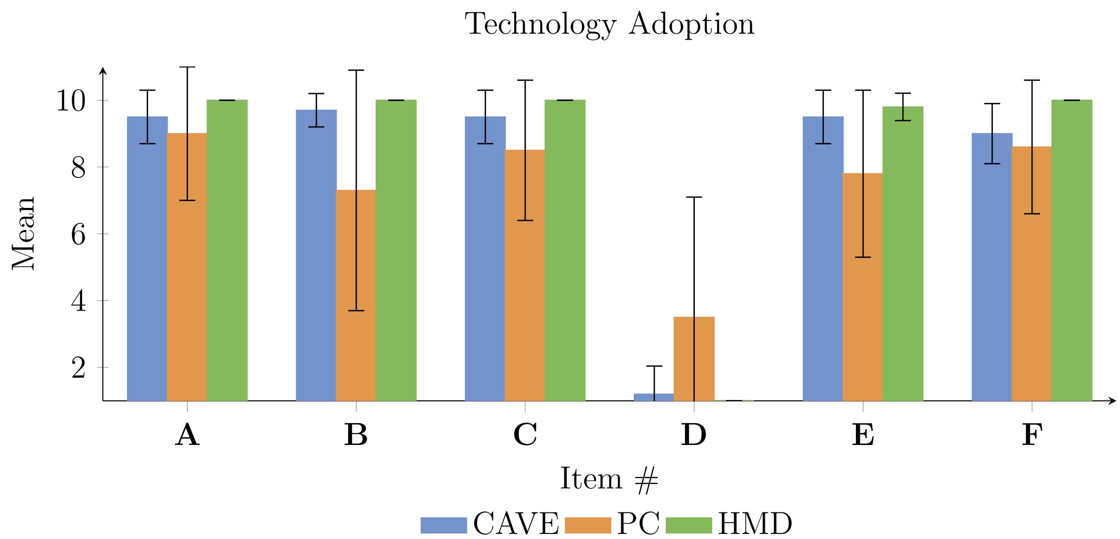


Figure 4.8: METIS UX: Technology Adoption Ratings. Mean and standard deviation of the Technology questionnaire (CAVE, n=6; PC, n=6; HMD, n=6). Items: **A**- If I use again the same virtual environment, my interaction with the environment would be clear and understandable for me; **B**- It would be easy for me to become skillful at using the virtual environment; **C**- Learning to operate the virtual environment would be easy for me; **D**- Using the interaction devices (Virtual reality headset, CAVE, controller and/or mouse) is a bad idea; **E**- The interaction devices would make learning more interesting; **F**- I would like learning with the interaction devices;

$SD = 0.83$; PC: $M = 8.5$, $SD = 2.07$; HMD: $M = 10$, $SD = 0$). For the PC condition, on the item: "Using the interaction devices (Virtual reality headset, CAVE, controller and/or mouse) is a bad idea" four participants considered that the use of the mouse was a bad idea, and two considered it was not a bad idea (PC: $M = 3.5$, $SD = 3.56$). On the other hand, the CAVE and HMD condition scored very low on this item (CAVE: $M = 1.17$, $SD = 0.40$; HMD: $M = 1$, $SD = 0$). Except for this item, overall participants agreed with the statements proposed in the Technology Adoption questionnaire (Figure 4.8).

UX: Experience Consequence

The experience consequence measures the perceived negative effect of the technology. Figure 4.9 presents the average and the standard deviation of all the items composing the subscale. Items are presented as statement, e.g. *"I suffered from vertigo during my interaction with the virtual environment"*, and participants need to indicate how much they agree with each item (Strongly disagree = 1; Strongly agree = 10). Overall, participants of the three conditions rated very low the negative effect of the technology on their wellbeing (CAVE: $M = 1.31$, $SD = 0.44$; PC: $M = 1.31$, $SD = 0.52$; HMD: $M = 1.06$, $SD = 0.11$).

UX: Users' Emotions

The results of the users' emotion questionnaire are presented in Figure 4.10. The questions of each emotion category where averaged to provide a mean score per emotion. Participants indicated, on a scale from 1 to 10, how much they agreed with 3 different statements for each emotion, e.g. *"I enjoyed being in this virtual environment"* for Enjoyment; *"I felt like distracting myself in order to reduce my anxiety"* for Anxiety; *"While using the interaction devices I felt like time was dragging"*

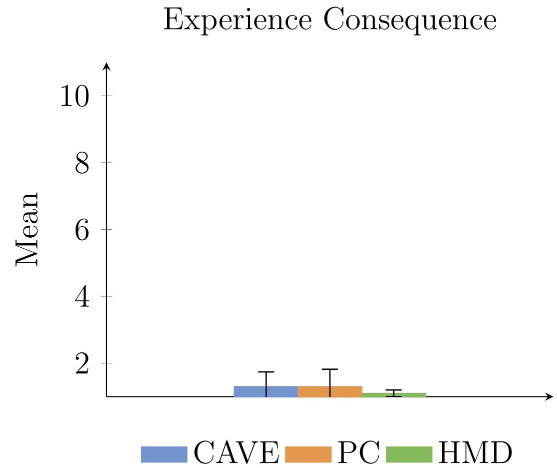


Figure 4.9: METIS UX: Experience Consequence Ratings. Mean and standard deviation of the Experience Consequence questionnaire (CAVE, n=6; PC, n=6; HMD, n=6). Items the subscale are averaged to compute an overall mean score.

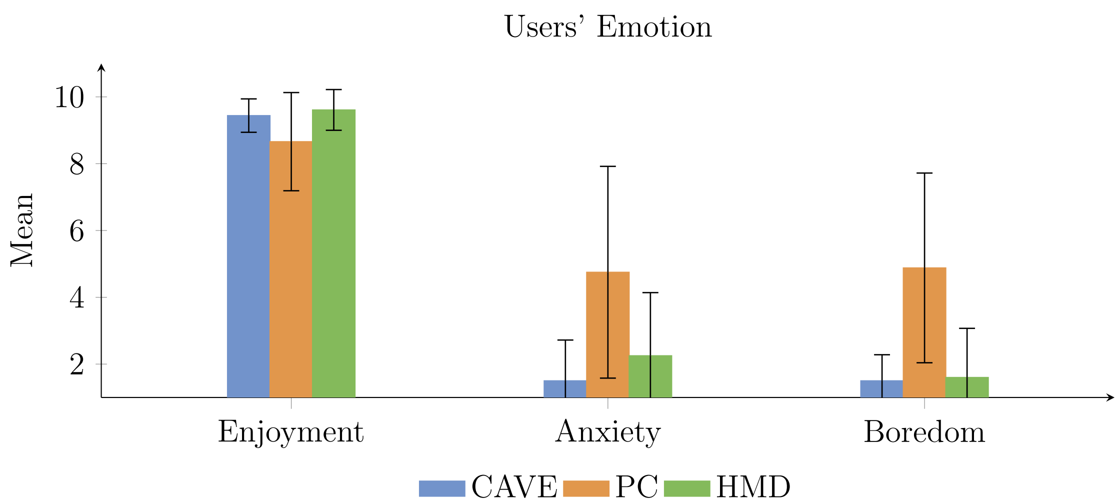


Figure 4.10: METIS UX: Users' Emotions Ratings. Overall mean and standard deviation of the Users' Emotion questionnaire (CAVE, n=6; PC, n=6; HMD, n=6). Items for each emotions are averaged to compute mean score per emotion (Positive activating: Enjoyment; Negative activating: Anxiety; Negative deactivating: Boredom.

for Boredom. Results indicate that all participants enjoyed interacting with the system, however it is not a clear cut for Anxiety and Boredom. As the HMD and the CAVE conditions scored low for both these emotions (CAVE: $M = 1.5$, $SD = 1.22$; HMD: $M = 2.25$, $SD = 1.89$ for Anxiety; CAVE: $M = 1.5$, $SD = 0.78$; HMD: $M = 1.66$, $SD = 1.47$ for Boredom), a higher average was observed for the PC (PC: $M = 4.75$, $SD = 3.17$ for Anxiety; PC: $M = 4.89$, $SD = 2.84$ for Boredom). The high standard deviation for both these emotions indicates that differences between the participants of the PC condition were also higher than in the other groups.

UX: Judgment

The judgment questionnaire measures the attractiveness of the system. Items such as *"I found that this virtual environment was Lame/Exciting"*, *"I found that this virtual environment was Amateurish/Professional"*, or *"I found that this virtual environment was Ugly/Beautiful"* were rated from 1 to 10, 1 representing the negative qualifying adjective. Results are presented in Figure 4.11. Overall, participants considered METIS as an attractive system for each platform (CAVE: $M = 9.33$, $SD = 0.52$; PC: $M = 7.87$, $SD = 1.53$; HMD: $M = 9.77$, $SD = 0.48$).

4.3.3 Discussion

The goal of this study was to assess the usability and the UX of METIS, a cross-platform classroom simulator. METIS is easy to use and easy to learn across platform and users enjoyed interacting with it (Figure 4.10). However, the CTA uncovered design shortcomings affecting users' interaction with METIS. This can indicate that the QUIS measure of usability was not sensitive enough to reflect these shortcomings. Nevertheless, the usability evaluation methods provided great insights in current

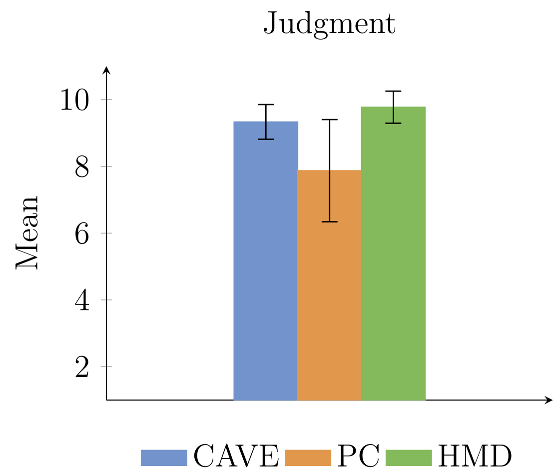


Figure 4.11: METIS UX: Judgment Ratings. Overall mean and standard deviation of the Judgment questionnaire (CAVE, n=6; PC, n=6; HMD, n=6). Items the subscale are averaged to compute an overall mean score.

limitations of the system. As the CTA helped detect usability issues shared across the platforms and issues specific to each, the semi-structured interviews helped refine their implications on the users and collect suggestions to improve the system.

Usability issues

The first issue observed was the identification of the disruptive students. Participants were able to accurately detect Jordan (100%) and Michael (89%), only one was able to correctly identify Ava (6%). Participants were caught off guard realizing which student was Ava, in the HMD and CAVE conditions, participants had to re-adjust their body position to face Ava. Errors in identifying Ava were caused by an inconsistency of the text introducing the situation and the behaviors of Ava in the classroom. The introduction specified that Ava's desk was cluttered and she was looking outside the window, while the virtual representation of Ava was looking to the board with only a few books on her desk. The look of the classroom and the initial behavior of the students were adapted to better reflect the introduction text.

A second usability issue was that some users (33%) had difficulty understanding that they had to interact with the system when reaching the first decision point in the simulation. A participant tried reading out loud one of the three options, then, when nothing happened, the participant noticed the laser and said; *"I don't have to repeat, I have to click"*. Another participant first reaction was: *"So I have to pick which one I want to say?"*. By trying different approaches, participants experiencing difficulties managed to make a selection. However, most participants realized they had to interact with the system. Moreover, 33% of the participants mentioned that they knew what to do at the first decision point because of the tutorial, illustrating the effectiveness of integrating a tutorial at the beginning of the simulation (*"Because of the instructions at the beginning, I know I have to make a choice."*). However, a participant noted the limitation of the tutorial which only shows one button panel. Based on this participant's suggestion the tutorial was modified to let users experience a selection of one panel between three options (as it appears during the simulation).

The CTA also revealed differences between the platforms. Whereas the laser pointer was well received with the HMD and the CAVE condition (a participant even used the laser to keep her position on text boxes), some PC users experienced difficulties with it. A participant qualified the laser on the PC as an: *"Interesting design choice"* and suggested that a *"normal clicker"* would be easier to manage. This comment was reflected by the Technology adoption results (Figure 4.8), three participants rated high the item "D - Using the interaction devices (Virtual reality headset, CAVE, controller and/or mouse) is a bad idea". The ambiguity of the question over the interaction devices being the overall system or just the controller/mouse lead us to think that some PC participants indicated their discomfort using the laser pointer with the PC platform. The laser pointer was replaced by a simple cursor for the METIS PC version.

The movement of the teacher with the CAVE platform caused some participants to experience dizziness, especially when the camera rotated to face the disruptive student. The rotation was implemented to be executed in a fixed time. For short rotation distances (happening frequently), the camera would slowly rotate to face the target position. However, in very few cases, the rotation distance can be really large thus causing fast movements of the camera. A CAVE participant said: *"That made me dizzy"* and suggested to slow down the rotation of the camera, as slower rotation did not have the same effect. Two CAVE participants experienced fast rotation in the CAVE and both reported it during the semi-structured interview. However, in the experience consequence survey (Figure 4.9) they only reported a mild negative effect. The rotations in the CAVE version were slowed down to prevent further negative effects on users. As the movements were implemented differently on the HMD platform (movements and rotation without an outside world reference can rapidly cause cybersickness), participants did not report any negative effects (Figure 4.9). One participant even said: *"The teleporting was a lot of fun!"*.

Realism

Many participants (66%) commented on the realism of the situations (*"It was pretty good at demonstrating those particular scenarios that are pretty common in a classroom"*), and on how they liked the proposed options (*"I appreciated the fact that there were the guiding options and most of the guiding options were realistic as towards what a teacher would actually do in that situation"*). However, 33% wanted more interactions with other students and the possibility to enter their own options.

As they agreed the METIS approach was a *"good first step"* in classroom behavior management training they also commented on its too simplistic approach:

- *"It is kind of simplistic. Often it is not a one on one relationship, it is a one to many. [..] You are interacting with one kid in the virtual world, however in the real world you might have two or three kids who are popping off."*
- *"Students are extremely aggressive. Felt like students were too calm. The body animation was fine, the voice should show the aggressiveness."*
- *"Sometimes it was a little too slow, in real classroom things will happen faster."*

A number of comments indicated that participants found METIS helpful:

- *"Nothing compares to being in an actual classroom but for the purpose of gaining experience and practicing, this kind of technology will really change. It has the potential to improve educators, being more comfortable in a classroom and getting to interact with student."*
- *"In class we might talk about what are the strategies but we don't experience it like in the simulations."*
- *"It feels more practice compared to the theoretical approaches I have in class. Real world practices always gonna be best in my opinion but, like, this is as close as it gets versus long lectures in classroom."*

Virtual student's display of emotions

The METIS virtual disruptive students were also equipped with the ability to display affective non-verbal behaviors using their body, hands, head direction and facial expressions in addition with body animations specified in the scenarios. However, only one participant reported on perceiving emotion from the virtual students' facial expressions. 55% percent indicated that they thought emotions were expressed by the students either from the body position, the audio speech or the context of the scenario. Some participants did not perceive the activated facial expressions, as evidenced by several suggestions to add facial expressions.

We have a few hypothesis on this observation. First, facial expressions portrayed may be too subtle to be perceived through the immersive VR. Another study explored immersive VR platforms for clinical expert training using a virtual patient able to display facial expressions [Ochs et al., 2018]. However, Ochs et al. did not comment on the perception of the facial expressions by users. Additionally, that study only used one virtual human, whereas participants using METIS might have been too distracted by other students in the classroom to perceive the main student protagonist's expressions. Secondly, the lack of eye-contact between the virtual student and the player could have hindered the observation of facial expressions of the student. In general, the virtual student face is directed towards objects (book, desk, phone, window) rather than towards the user trainee. A third hypothesis, is that the duration of display of the facial expression might have been too short for participants to detect them. A participant commented that the speech bubbles sometimes distracted her from observing the virtual student's actions.

To increase the perception of facial expressions, three modifications were implemented on the METIS system. First, the facial expressions duration was increased. Rather than displaying the facial expressions for a pre-defined period of time, the facial expression stays on until a different facial expressions or a reset is called by the system. Secondly, in the revised version, the students' head faces the teacher (the camera) every time the student is addressed or is addressing the teacher. Finally, the display of speech bubbles was modified to only display bubbles in the case the student actions could be ambiguous. This modification consisted of going through the scenarios and annotate which bubbles to display (based on the scenario context and on the quality of some animations). The number of speech bubbles displayed is reduced by 90% in the new version of METIS.

VR platforms comparison

When comparing different VR platforms for a given system, one must wonder which aspects can be compared and which do not. As some applications may be better suited for certain type of VR platforms, how to make sure that the comparison of the different platforms is fair if interaction designs are different. In the other hand, an interaction design shared across VR platforms might not exploit the full potential of each device.

How to know if the maximum potential for a platform have been reached? It is hard to answer that question, however user-centered design can bring ideas on how to improve for each single VR platform. For instance, when interacting with the HMD, some users tried to pick up books and give them to students, they also controlled their distance with students (physically moving away or closer given the situation). However, if one VR platform allows for more interaction than another with the same system, are the usability and UX evaluations of each VR platform even comparable? This raise the question of should a cross-platform system aim for equivalence of interaction design?

In this study, the same type of interactions and UI were implemented for all platforms. The main reason for this was imposed by the design of the METIS system. The METIS simulator follows pre-scripted vignette scenarios which only allow a limited number of possibilities to continue through the story. In the case of METIS, we showed that a similar interaction design shared across platform can achieve equivalent usability, thus laying the groundwork for an efficient use of the system and potentially efficient learning. We leave to future studies the usability comparison of a cross-platform system using different interaction for its different platforms.

We proposed an approach to evaluate the interaction design of cross platform VTEs and applied it on METIS to answer (**RQ3**). We established that METIS is easy to use and easy to learn across all three platforms. The technology adoption, experience consequence, and user's enjoyment questionnaires revealed promising results for METIS. A CTA was used to evaluate the system and has proven useful to collect feedback on the system and on individual platform. Feedback underlined useful features (tutorial) and preferred features for each platform (laser pointer vs. cursor). Finally, interviews showed that participants would like to use METIS in a learning context, which will be part of our future research.

4.4 Evaluation of METIS: VR Platforms and Emotional Students

The goal of this evaluation is to explore the effects of immersive VR and emotional virtual agents on user engagement, believability, perceived presence and co-presence (**RQ3, RQ4,** and **RQ5**). We considered three **VR Platforms**: **PC**, **HMD**, and **CAVE**. METIS serves as the application to compare the effect of our three VR platforms for two conditions: one where students can display facial expressions (**FE**) and one with only neutral expressions (**Neutral**). We established the following hypotheses:

- **H1 - HMD/CAVE vs PC**: The HMD and the CAVE will generate more UX than the PC.
- **H2 - FE**: There will be an effect of the display of facial expressions on UX for all platforms (PC, HMD, CAVE).
- **H3 - FE vs Platforms**: The display or not of facial expressions will influence UX depending on the VR platform used (Interaction effect).
- **H4 - HMD vs CAVE**: There will be no difference of impact on UX between the HMD and the CAVE.

For the second hypothesis, **H2 - FE**, we expect to see an impact of facial expressions for each VR platform on UX, however, as observed by Beale et al. for different studies [Beale and Creed, 2009], this impact could be either detrimental or beneficial.

The results presented in this section were part of published work [Delamarre et al., 2020a].

4.4.1 Methodology

In this section, we first present the experimental design, followed by measures and materials used. We then describe the experiment protocol and give details about the population recruited for this study.

Design

As mentioned, the goal of this study is to explore the interactions between VR platforms, and the display of facial expressions by virtual agents and their effect on presence, co-presence, engagement, and believability. We established three VR platforms, the traditional desktop setup (PC), to which we compared a VR headset (HMD) and a virtual room (CAVE). Participants of each group interacted with two versions of the METIS classroom simulator, one in which the disruptive virtual students exhibited facial expressions (FE) and one in which the students' faces remained neutral (Neutral).

For each measure we consider three effects: (1) effect of the display of facial expressions across all platforms (within effect, Neutral vs. FE); (2) effect of both the platform and the display of facial expressions on the measure (interaction effect); and (3) effect of the platforms on the measure (between effect, CAVE vs. PC vs HMD).

Population

A total of 63 participants were recruited from a variety of graduate and undergraduate programs of a U.S. public university. An outlier removal phase helped identify 5 outliers. We considered outliers participants who entered the same score to every items of a questionnaire every time this questionnaire was completed (for both with and without facial expressions conditions) and participants for which a questionnaire score fell outside the inner Tukey's fences ($1.5 \times IQR$, IQR: Interquartile range). Demographics about the remaining 58 participants are detailed in Table 4.1.

Procedure

The flow of the experiment is showed in Figure 4.12. First, participants were given information about the nature of the study through a consent form approved by Florida International University's Internal Review Board, and were offered the opportunity to ask questions about their participation. The experimenter reminded participants on multiple occasions that they could stop the experiment and withdraw

		Group		
		PC	HMD	CAVE
Total		20	19	19
Gender	Female	8	8	7
	Male	12	11	12
VR experience	HMD	13	12	10
	CAVE	6	6	5
University education program	Education	2	3	5
	Comp. Sci.	8	8	4
	Other	10	8	10

Table 4.1: Number of participants for each group and gender, participants who already experienced a HMD or CAVE platform and education program

their consent at any time for any reason (e.g. feeling uncomfortable with the content, or dizzy because of the technology).

Once the consent form was signed, the experimenter assigned participants to a VR platform randomly. Then, participants started the first set of three scenarios. The order of interaction with the two conditions (with and without facial expressions) was randomized to counter the order effect. Each set of three scenarios started with a quick tutorial (described in Section 4.1.2) embedded in the simulation teaching participants how to interact with their assigned platform. After completing the first set of scenarios, participants completed the presence and co-presence questionnaires. Next, participants completed the second set of three scenarios and following the questionnaires. Before leaving, participants filled a demographics questionnaire and then received a print out of their feedback.

Measures

In order to measure the presence perceived by participants, we adapted a self-report questionnaire designed by Bailenson et al. [Bailenson and Yee, 2006]. The goal of

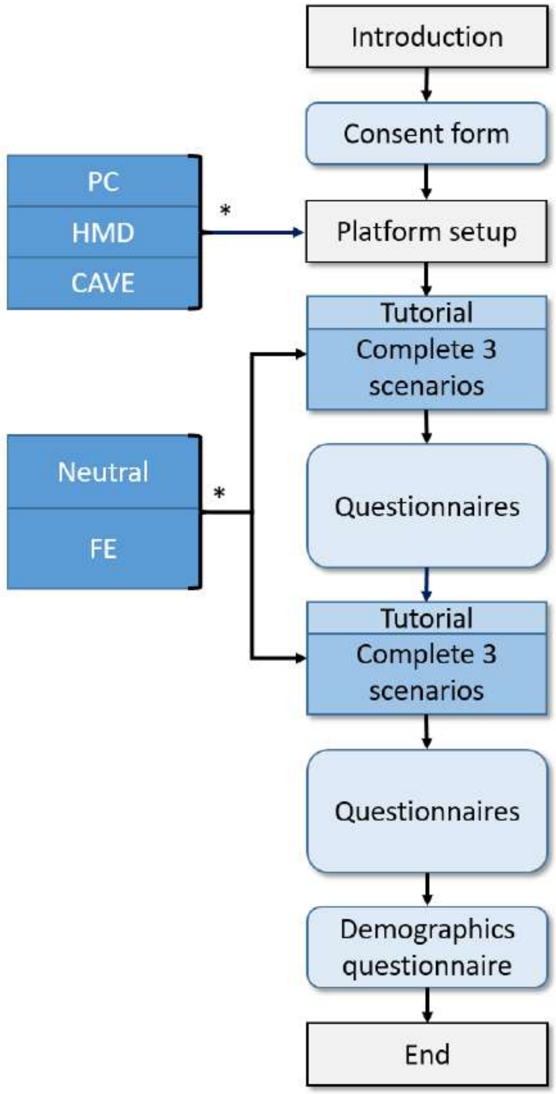


Figure 4.12: METIS VR Platforms and Emotional Student Experiment Procedure. The * indicates were a random selection occurred (e.g. platform and order of condition). Participants interacted with 3 classroom scenarios for each condition (Neutral and with facial expressions - FE).

the questionnaire is to measure the feeling of "being there" in the environment and is composed of four items, each item rated on a scale from 1 to 7.

Similarly, the perceived co-presence was adapted from a questionnaire developed by Bailenson et al. [Bailenson and Yee, 2006]. This questionnaire is designed to evaluate the human-likeness and the sociability of the virtual agents. It contains five items rated on a scale from 1 to 7.

Engagement was measured using the UES-SF (Short Form), a questionnaire composed of 12 items divided into 4 categories: aesthetics appeal, focused attention, perceived usability and reward [O'Brien et al., 2018]. Each item is rated on a scale from 1 to 5.

Believability was measured using a self-report questionnaire designed by Gomes et al. which includes nine items based on dimensions of believability, each item rated on a scale from 1 to 7 [Gomes et al., 2013] : (1) Awareness; (2) Behavior understandability; (3) Personality; (4) Visual Impact; (5) Predictability; (6) Behavior coherence; (7) Change with experience; (8) Social; and (9) Emotional expressiveness. High values for each dimension result in a higher sense of believability with the exception of predictability, where too much predictability or too little can be detrimental for believability.

Finally, participants indicated how they perceived the emotions of each student (Ava, Jordan, Michael) by selecting one or many of the following options: *Body Language, Vocal intonation, Choice of word,Facial expressions*, and *None*.

Apparatus

For the PC and HMD conditions, the classroom simulator was launched on a Corsair 64bit with 16GB of RAM, an Intel processor CPU i7-4790k 4.00GHz and a NVIDIA GeForce GTX 1080 Ti graphics card. The operating system was Windows 10 Education. The HMD used is a HTC Vive with its respective controllers. The CAVE is a 5-wall hexagon shaped (5.7m $\times$ 6m). It includes real-time head and controller tracking and active stereoscopic display.

4.4.2 Results

Before performing the statistical analysis, we looked at the Cronbach's alpha of each questionnaire for both conditions (FE and Neutral). We removed question items only when the Cronbach's alpha was improved for both conditions.

For the presence questionnaire, we observed that removing question 2, just as was found in [Bailenson and Yee, 2006], increased the Cronbach's alpha of both conditions (FE:from $alpha = .49$ to $alpha = .5$; Neutral: from $alpha = .62$ to $alpha = .66$). Similarly, for the co-presence questionnaire, removing item 3 increased the Cronbach's alpha for both condition (FE: from $alpha = .83$ to $alpha = .85$; Neutral: $alpha = .76$ to $alpha = .83$). The Cronbach's alpha of the engagement questionnaire could not be improved for both conditions, no item was removed from the engagement questionnaire (FE: $alpha = .84$; Neutral: $alpha = .87$) Finally, the believability questionnaire Cronbach's alpha was improved by removing the predictability question item (FE: from $alpha = .60$ to $alpha = .72$; Neutral: $alpha = .55$ to $alpha = .66$).

Presence

Self-reported presence scores were analyzed using a 3×2 mixed design ANOVA for the three VR platforms (PC,HMD,CAVE) and the two conditions (Neutral, FE), in which the VR platforms served as the between-subjects variable, and the display or not of facial expressions served as the within-subjects variable. Mean presence score for the VR platforms and for both conditions (Neutral, FE) are shown in Figure 4.13. The mean presence scores ranged from $M = 4.48$ (PC,$SD = .87$) to $M = 5.29$ (HMD,$SD = 1.27$) for the Neutral condition and from $M = 4.80$ (PC,$SD = 1.05$) to $M = 5.17$ (CAVE,$SD = 1.16$) for the FE condition.

No difference were observed for presence between the Neutral and the FE conditions (within effect: $F(1, 55) = .479$, $p = .492$). No interaction effect between VR platforms and the display of facial expressions was observed for presence (interaction effect: $F(2,55) = 2.341$, $p = .106$). There was no difference in presence between the VR platforms (between effect: $F(2,55) = 1.578$, $p = .216$).

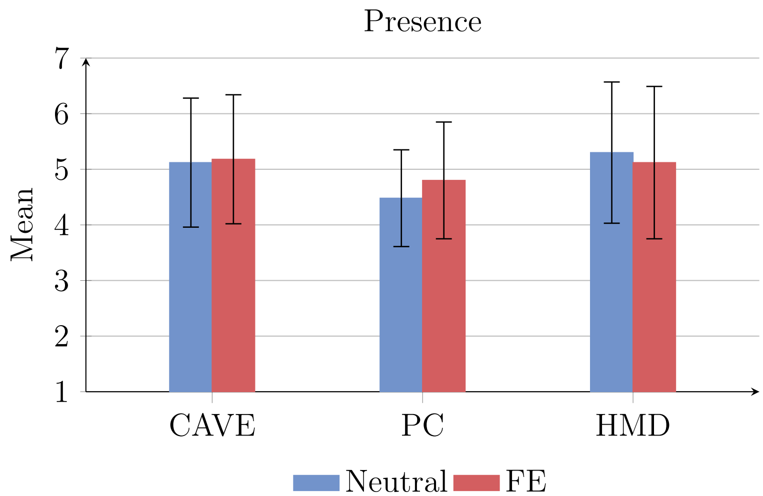


Figure 4.13: METIS: Presence Ratings. Overall mean and standard deviation of Presence questionnaire responses for all VR platforms (CAVE,PC,HMD) and for the two conditions Neutral and FE. The presence questionnaire is composed of three items, each item rated on a scale from 1 to 7

Co-presence

Self-reported co-presence scores were analyzed using a 3×2 mixed design ANOVA for the three VR platforms (PC,HMD,CAVE) and the two conditions (Neutral, FE), in which the VR platforms served as the between-subjects variable, and the display or not of facial expressions served as the within-subjects variable. Mean presence score for the VR platforms and for both conditions (Neutral, FE) are shown in Figure 4.14. The mean co-presence scores ranged from $M = 4.27$ (PC,$SD = 1.51$) to $M = 4.92$ (CAVE,$SD = 1.30$) for the Neutral condition and from $M = 4.48$ (HMD,$SD = 1.30$) to $M = 4.78$ (CAVE,$SD = 1.51$) for the FE condition

No difference were observed for co-presence between the Neutral and the FE conditions (within effect: $F(1,55) = .793$, $p = .377$). No interaction effect between VR platforms and the display of facial expressions was observed for co-presence (interaction effect: $F(2,55) = 2.495$, $p = .092$). There was no difference in co-presence between the VR platforms (between effect: $F(2,55) = .562$, $p = .574$).

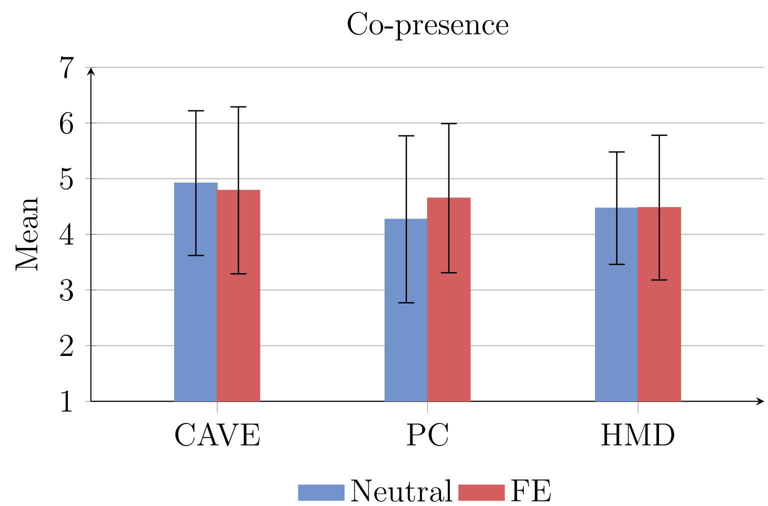


Figure 4.14: METIS: Co-presence Ratings. Overall mean and standard deviation of co-presence questionnaire responses for all VR platforms (CAVE,PC,HMD) and for the two conditions Neutral and FE. The co-presence questionnaire is composed of four items, each item rated on a scale from 1 to 7.

Engagement

Self-reported engagement scores were analyzed using a 3×2 mixed design ANOVA for the three VR platforms (PC,HMD,CAVE) and the two conditions (Neutral, FE), in which the VR platforms served as the between-subjects variable, and the display or not of facial expressions served as the within-subjects variable. Mean engagement score for the VR platforms and for both conditions (Neutral, FE) are shown in Figure 4.15. The mean engagement scores ranged from $M = 4.05$ (HMD,$SD = .52$)

to $M = 4.17$ (CAVE,$SD = .58$) for the Neutral condition and from $M = 4.03$ (PC,$SD = .49$) to $M = 4.20$ (CAVE,$SD = .51$) for the FE condition

No difference were observed for engagement between the Neutral and the FE conditions (within effect: $F(1, 55) = .643$, $p = .426$). No interaction effect between VR platforms and the display of facial expressions was observed for engagement (interaction effect: $F(2, 55) = .018$, $p = .982$). There was no difference in engagement between the VR platforms (between effect: $F(2, 55) = .572$, $p = .568$).

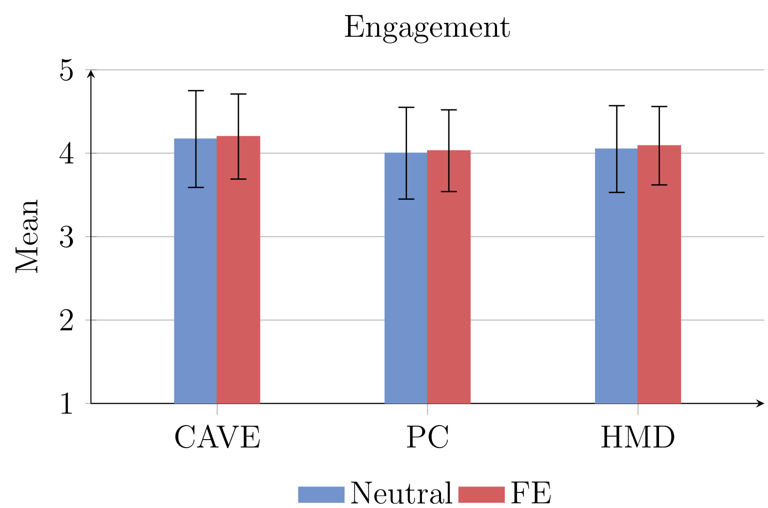


Figure 4.15: METIS: Engagement Ratings. Overall mean and standard deviation of engagement questionnaire responses for all VR platforms (CAVE,PC,HMD) and for the two conditions Neutral and FE. The engagement questionnaire is composed of 11 items, each item rated on a scale from 1 to 5.

Believability

Self-reported believability scores were analyzed using a 3×2 mixed design ANOVA for the three VR platforms (PC,HMD,CAVE) and the two conditions (Neutral, FE), in which the VR platforms served as the between-subjects variable, and the display or not of facial expressions served as the within-subjects variable. Mean believability score for the VR platforms and for both conditions (Neutral, FE) are shown in

Figure 4.16. The mean believability scores ranged from $M = 5.34$ (HMD,$SD = .77$) to $M = 5.40$ (CAVE,$SD = .86$) for the Neutral condition and from $M = 5.45$ (PC,$SD = .59$) to $M = 5.51$ (HMD,$SD = .64$) for the FE condition

No difference were observed for believability between the Neutral and the FE conditions (within effect: $F(1, 55) = 2.518$, $p = .118$). No interaction effect between VR platforms and the display of facial expressions was observed for believability (interaction effect: $F(2, 55) = .135$, $p = .874$). There was no difference in believability between the VR platforms (between effect: $F(2, 55) = .020$, $p = .980$).

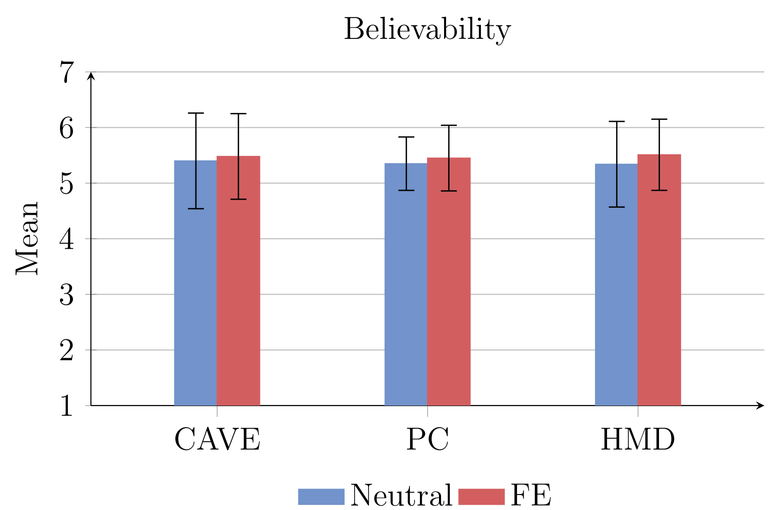
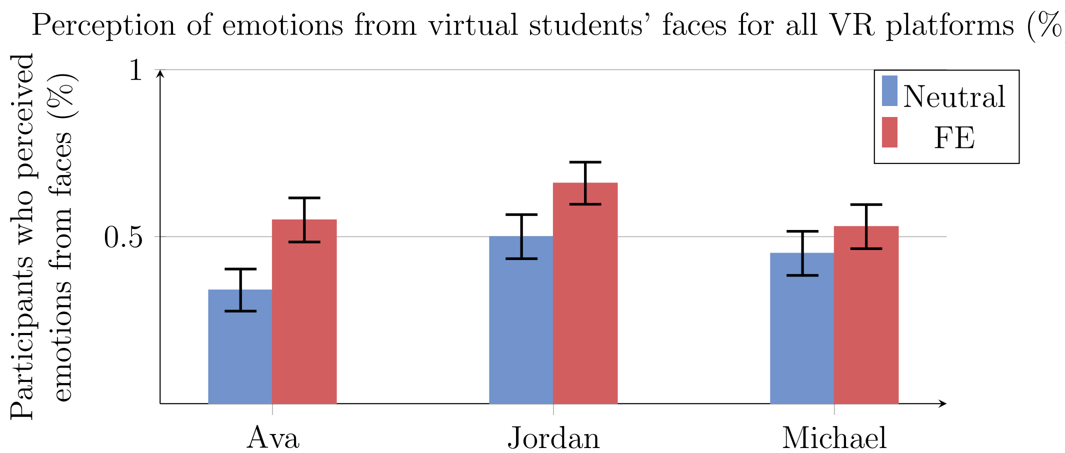


Figure 4.16: METIS: Believability Ratings. Overall mean and standard deviation of believability questionnaire responses for all VR platforms (CAVE,PC,HMD) and for the two conditions Neutral and FE. The believability questionnaire is composed of seven items, each item rated on a scale from 1 to 7

Perception of Facial Expressions

For each condition (Neutral and FE) and for each student (Ava, Jordan, Michael) participants indicated if and how they perceived emotions from the student by selecting one or many of the following options: *Body Language, Vocal intonation, Choice of word, Facial expressions*, and *None*. We conducted a McNemar's Chi-



Ava *($p = .004 < .05$)		Jordan *($p = .033 < .05$)		Michael ($p = .150 > .05$)	
 Neutral	 Embarrassment	 Neutral	 Anger	 Neutral	 Sadness

Table 4.2: METIS: Facial Expressions Perception. Means and SE of answers indicating that emotions expressed by the virtual students were perceived from the faces for Neutral and FE for all platforms. Neutral faces and examples of facial expressions are shown for each student.

Square test [Fisher et al., 2011] for each category (0 if not selected, 1 if selected) comparing between the Neutral and the FE conditions. No differences were observed for *Body Language, Vocal intonation, Choice of word*, and *None*.

As shown in Figure 4.2, differences were observed for Ava and Jordan regarding the perception of emotions from the students' faces ($p = .004 < .05$ for Ava, and $p = .033 < .05$ for Jordan). No significant difference was observed for Michael between Neutral and FE conditions ($p = .150 > .05$).

4.4.3 Discussion

Our results show that presence was rated above average for each VR platform (PC, HMD, CAVE) and each condition (Neutral vs FE), indicating that participants interacting with METIS had a feeling of being in an actual classroom rather than a computer science lab. Similarly, co-presence scores show that for each VR platform participants felt as though they were with someone in the virtual classroom. For all three VR platforms and both conditions, participants felt highly engaged with the METIS simulator (the engagement questionnaire averages at 4.10 for METIS; an existing educational serious game engagement scores averages at 3.71 [Andrade and Law, 2018]). Finally, results show that participants were able to suspend their disbelief when interacting with the students for all VR platforms (believability was rated above average 5.42).

No difference could be observed for all UX measures between the VR platforms. It means that participants felt similar level of presence, co-presence, engagement, and believability regardless if they were seated at a desk in front of a screen, or surrounded by a virtual environment. These results contradict studies which observed more presence and engagement for the immersive VR platforms [Buttussi and Chittaro, 2018, Ochs et al., 2018]. However, in our study, participants only interacted with one platform, therefore they could not form a comparative opinion between platforms. This show that the PC version, in itself, is able to generate positive feeling of presence, co-presence, believability and engagement. The between-subjects study design was chosen to increase experimental controls and reduce confounding variables [Cao et al., 2019, McMahan et al., 2012].

However, several limitations of this study can explain why we did not observed increases in UX for the immersive VR platforms. First of all, our sample size (n=58) limits our ability to detect significant differences to observations with large effect sizes. The subjective nature of measures such as presence, co-presence, and believability and the subtle changes generated by facial expressions may require a larger sample size. A second limitation lies in the fact that our population is not entirely composed of teachers or students in the process of becoming teachers. Ochs et al. observed that participants from the domain showed more involvement with the application compared to "naive" participants [Ochs et al., 2018]. Finally, the Cronbach's alpha of our presence questionnaire was low as observed in other studies [Bailenson and Yee, 2006]. Other presence questionnaires exists such as the Igroup Presence Questionnaire (IPQ) [Schubert et al., 2001] composed of 14 items could help validate our first set of results.

As no observation could be made when comparing the three VR platforms at once, we explored our results by doing pairwise comparisons (PC vs. CAVE, PC vs. HMD, HMD vs. CAVE) and by comparing the effect of facial expressions (Neutral vs FE) within each VR platform.

For this exploration phase, no statistical adjustment was used. Statistical adjustment for multiple independent comparisons minimizes type 1 error (false-positive) but increases type 2 error (false-negative) [Rothman, 1990, Perneger, 1998]. In most cases, minimizing type 1 error is essential to prevent validating a hypothesis when the results can be attributed to chance. However, the goal of the exploration phase is to lay the groundwork for future work in order to study the use of non-verbal behaviors with immersive VR technology.

Each measure (presence, co-presence, engagement, believability) were submitted to 2×2 mixed design ANOVA for each comparison (PC vs. CAVE, PC vs. HMD, HMD vs. CAVE), in which the VR platforms served as the between-subjects variable, and the presence or absence of facial expressions served as the within-subjects variable. Again, no differences were observed for engagement and believability.

However, interaction effects between the VR platforms and the facial expressions were observed for presence and co-presence.

Exploring the Effect of VR Platforms and Facial Expressions on Presence

The ANOVA for the PC $\times$ HMD comparison revealed that there was an interaction effect between the VR platforms and facial expressions on presence, $F(1, 37) = 5.249$, $p = .028 < .05$. We can observe that presence scores for the PC are higher when virtual students display facial expressions (FE: $M = 4.80$, $SD = 1.05$) than when no facial expressions are displayed (Neutral: $M = 4.48$, $SD = 0.87$). For the HMD, the inverse trend is observed (FE: $M = 5.12$, $SD = 1.37$; Neutral: $M = 5.30$, $SD = 1.27$).

A follow up t-test revealed that HMD presence scores for the Neutral condition were significantly higher that PC Neutral presence scores ($p = .027 < .05$). No difference was observed for the FE condition.

Presence scores were not different for the other two comparisons (PC $\times$ CAVE and HMD $\times$ CAVE).

Exploring the Effect of VR Platforms and Facial Expressions on Co-presence

An interaction effect between the VR platforms and facial expressions was observed for the PC $\times$ CAVE comparison ($F(1, 37) = 6.073$, $p = .018 < .05$). Co-presence scores increase for the PC when the students display facial expressions (FE: $M = 4.65$, $SD = 1.34$; Neutral:$M = 4.27$, $SD = 1.51$). On the other hand, for the CAVE, FE co-presence scores (FE:$M = 4.78$, $SD = 1.51$) are lower than Neutral co-presence (Neutral: $M = 4.92$, $SD = 1.30$).

A follow up t-test revealed more co-presence for the FE condition compared to the Neutral condition for the PC ($p = .027 < .05$). No difference was observed for the CAVE.

Co-presence scores did not reveal differences for the other two comparisons (PC × HMD and HMD × CAVE).

Conclusion

The exploration phase revealed two observations. First, we observed a difference in terms of feeling of presence between the HMD and the PC. A similar tendency for presence was also observed between the CAVE and the PC ($p = .058 > .05$). These observations reflect results of existing studies comparing presence between desktop and immersive platform [Buttussi and Chittaro, 2018, Ochs et al., 2018]. Therefore, even though we reject **H1 - HMD/CAVE vs PC** for engagement, believability and co-presence, we do not reject **H1** for presence.

Secondly, the exploration revealed that FE co-presence scores for the PC were higher that neutral co-presence scores, indicating that the display of facial expressions with the PC version increases the feeling of co-presence. This observation replicates findings from a previous study on the effect of avatars in virtual environments [Casanueva and Blake, 2001]. As we reject **H2 - FE** for the CAVE and the HMD platform, we cannot reject **H2** for the PC.

These two observations lead to an interaction effect between the HMD and the PC for presence and between the CAVE and the PC for co-presence, therefore we accept **H3 - FE vs Platforms**. As no difference was observed between the HMD and the CAVE for all measures we also accept **H4 - HMD vs CAVE**.

Participants reported perceiving emotions from facial expressions more for two students (Ava and Jordan) on all platforms. They also reported perceiving emotions from facial expressions when no facial expressions were present. Ragan et al. observed that virtual environment with higher levels of details worsened performance for visual scanning tasks with immersive VR [Ragan et al., 2015]. This observation could explain why facial expressions had less of an impact in the HMD and the CAVE. The large peripheral view provided by the immersive VR platforms (participants can see more detail of the classroom and more students) may have lowered the impact of facial expressions on participants' feeling of presence and co-presence. Additionally, when participants are looking at the disruptive students with the CAVE and the HMD they can still see other students which do not display facial expressions. With the PC, the disruptive student is always at the center of the screen and therefore the attention of the participants are directed mostly on this student. In future work we propose a study to compare the perception of facial expressions with different VR platforms.

Nevertheless, all our virtual students use the same rig and the same animations. The difference of perception of emotions between the virtual students (Ava, Michael, Jordan) is the results of participants' subjective observations. We hypothesize that participants may have had greater issues identifying the facial expressions for Michael because of lighting and contrast. Moreover, previous research showed that users prefer interacting with virtual humans of the same ethnicity [Cowell and Stanney, 2005, Baylor et al., 2003, Moreno and Flowerday, 2006]. The fact that the majority of our participants were white (60%) could also explain this observation.

CHAPTER 5

FUTURE WORK

The goal of VTEs for social skills is to recreate realistic situations to provide efficient ways of training. This is not an easy task as it involves many aspects such as the creation of realistic graphics, situations, and interactions. METIS, by respecting the proposed requirements successfully addressed these challenges and offer a test platform for future studies. We present the next step of METIS in the following sections.

5.1 Effect of facial expressions on VR

The exploration on the effect of facial expressions with immersive VR platforms revealed that facial expressions have no effect with HMD and the CAVE. However, as underlined by previous studies, the more details that are visible the harder it is for users to notice specific features [Ragan et al., 2015]. To study the perception of facial expressions using immersive VR platforms future studies should compare two conditions across the METIS VR platforms. The first conditions would include the classroom used in this study, full of students and clutters on the desk, while the second condition would contain no clutters and only one virtual student. The disruptive student must be the same between the two conditions to ensure a sound comparison and the population used must include equal proportions of different ethnicity to cover potential race and ethnicity effects of the human-virtual human interactions [Moreno and Flowerday, 2006].

5.2 Affective Autonomous Model for Virtual Students

The current approach used by METIS regarding the display of affective behaviors relies on pre-scripted scenarios using pre-scripted display of emotions (i.e. at each moment the scenario controls with non-verbal behaviors students are showing). This approach allows the author of the vignette to control every aspect shown to users, however as scenarios become larger this approach becomes more challenging to manage. Additionally, as we mentioned in our *Scalability* requirement, the variability of the situation presented to users allow them to abstract knowledge and benefits learning [Bossard et al., 2008]. A pre-scripted approach requires the generation a lot of content to ensure variability.

To overcome these limitations, a potential approach is to resort to a model of emotions. In the last decade researchers have focused on creating virtual emotional entities that are able to understand and express emotion [Lisetti et al., 2013]. In order to generate accurate emotions and behaviors, as an actual human would do in a similar situation, affective computing researchers, basing their work on psychological theories of emotion [Ortony et al., 1988, Scherer, 2009, Lazarus and Lazarus, 1991], attempted to model the mechanisms behind emotion generation [Gratch and Marsella, 2004, Dias et al., 2014, Becker-Asano, 2014].

To address this challenge, we propose the Appraisal Interpersonal Model of Emotion Regulation (*AIMER*), an emotion-based architecture to enable the generation of autonomous socially adapted behaviors that are non-repetitive [Delamarre et al., 2019a] (Figure 5.1). Our approach to design our affective model will be done in three steps:

1. Emotion generation process - Based on events taking place in the virtual classroom, our model will determine the virtual student's emotional states. The first step is to determine which emotions we want for our virtual students, we will start with the seven emotions identified for METIS 4.6. Secondly, we will consider Scherer's appraisal theory of emotions to determine the appraisal variables necessary to generate these emotions [Scherer, 2009]. Scherer proposes a multilevel sequential approach for the appraisal process. An event will be appraised sequentially by four stimulus evaluation checks (SEC): (1) relevance; (2) implication for self and others; (3) coping potential; and (4) normative significance. Each SEC treat the input information and passes on to the next check (hence the sequential aspect). Moreover, the SECs occur on all three levels of the emotion processing system proposed by Leventhal [Leventhal, 1984]: (1) Sensory motor level; (2) Schematic level; and (3) Conceptual level (hence the multilevel aspect). Lower levels are usually faster to determine a SEC, however the appraisal information is can be inaccurate. For example a sudden event can trigger surprise, then fear and then relief expressed by a laugh, surprise and fear were generated by the physiological response to an event without having processed all the information, realizing that the event is not harmful, relief is then generated. A benefit of Scherer's approach is that it generates emotion during the appraisal process thus representing the ephemeral aspects of emotions. Finally, we need to link events happening in the environment to the SECs. Events in the environment are triggered by a teacher action which uses emotion regulation strategy. Therefore, to model the relationship between the virtual disruptive student and the environment we will use Interpersonal Emotion Regulation theory (IER)[Niven et al., 2009]. Niven's classification of interpersonal emotion regulation strategies [Niven et al., 2009] will be used to categorize the teacher actions. Niven proposes two types of strategies: Affect-improving strategies such as positive engagement, humor, distraction; and Affect-worsening strategies like negative engagement, criticizing, showing disrespect. For each strategies, Niven also provide action prototypes (Criticizing: *"Pointing out the target's flaws"*; Distraction: *"Arranging social activity for the target"*). By mapping IER strategies to Scherer's SECs, from a emotion regulation strategy applied by the teacher trainee, the model is able to generate a sequence of emotions relevant to the situation.

2. Non-verbal behavior generation - In order to generate realistic non-verbal behaviors such as gaze, head, and body movements, given a virtual student emotional state, we will analyze freely accessible videos of children displaying emotions. Each video will be tagged with their corresponding emotions, head and body movements. We will then generate a 3D virtual representation. Body and head movements will be recorded using motion capture by mimicking children behaviors and facial expression will be represented using METIS's facial expressions.

The virtual representation will then be iteratively evaluated by a board of experienced teachers using questionnaires and focus group and refined accordingly until the virtual representation of the emotion is deemed believable. The addition of non-verbal behavior will enhance the affective believability of the interactions with the virtual students as non-verbal behaviors will be dynamically adapted to the teachers action (e.g. if the teacher decide to take one student's phone away, the student's neutral expression will change to an angry one, taking an aggressive posture and furiously starring at the teacher). We will use BML [Kopp et al., 2006] to command the virtual students behaviors.

3. Action generation - Building on the emotion generation process described in the first step, once a emotion is generated for a virtual student we want to determine an action relevant to its goals and affective internal states. Gross proposes five emotion regulation strategies [Gross, 2015]: (1) Situation selection, i.e. influencing the situation to be exposed to; (2) Situation modification, i.e. modifying the situation; (3) Attentional deployment, i.e. focusing or ignoring parts of the situations; (4)

cognitive change), i.e. changing the cognitive representation of the situation; and (5) response modulation, i.e. modifying an emotion-related actions. For a given appraised event, the student will sequentially: (1) identify a situation (internal states, social relationship, peer student vs teacher); (2) select a strategy to apply (among the different strategies proposed by the model and existing in the vignettes); and (3) finally apply that strategy. Each strategy will be represented by a set of potential actions. Virtual students actions are constituted of two components: (a) Operations (e.g. taking an object, moving to a position); (b) Utterances (e.g. verbally addressing the teacher or other virtual students). The action will be selected from the ones already implemented in the METIS system (approx. 80 actions).

For a given teacher actions towards the disruptive student (e.g. confiscating the phone), first an emotion will be determined (e.g. anger), which will then generate non-verbal behaviors (e.g. angry expressions, closing hands and, starring at teacher) and select an appropriate action (e.g. stand up and snatch the phone back).

By implementing *AIMER* within METIS, future work will be able to observe how virtual students controlled by a model of emotions can impact user engagement, suspension of disbelief and feeling of co-presence compared to pre-scripted approaches.

5.3 Immersive Interactions

By porting METIS to immersive VR, and more particularly to HMDs, a new set of interactions with the virtual classroom becomes possible. During our studies of METIS, participants using the HMD tried to interact with the virtual environment: many tried to grab a book to give it to the student, one participant was pacing while lecturing the classrooms, and another one tried to pat Michael's back when he was crying.

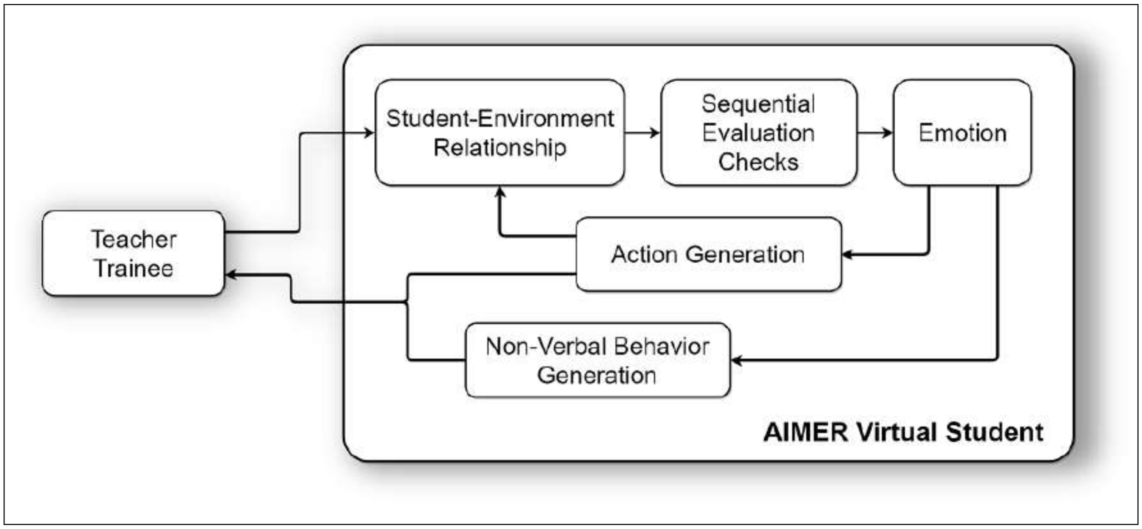


Figure 5.1: AIMER High-Level Architecture

Using METIS with the HMD, student teachers could learn about the appropriate distance to have with a student or when to initiate physical contact (e.g. hand on shoulder). By coordinating with education experts and experienced teachers, future work would be able to identify VR interactions which can benefit teachers' education.

Additionally, many users tried to speak out loud the different choices that were proposed by the system. Therefore, we plan to integrate speech recognition technologies to METIS in order to remove the need for a controller while interacting with the system. The recording of teacher trainees' voices will also be used to generate targeted feedback on intonation of the utterance (e.g. *in this context a firm tone would be more appropriate*) and volume of the voice (e.g. *given the overall classroom sound level, you should speak louder to make sure all the students can hear you*)

5.4 Vignette Authoring System

METIS currently uses scenarios created in LucidChart [LucidChart, 2020]. As LucidChart was the platform selected by pedagogical experts to design the classroom scenarios, the approach taken by METIS allows experts to use their own formalism for the content. This approach however requires experts to use a tagging system which, when not used correctly, generate errors in the simulation. Additionally, as the scenarios extended, inconsistencies appeared (e.g. asking us to program the teacher to move to a student desk, even though the teacher is already at the desk from a previous action in the scenario, or specifying to open book even though the book is already opened from an earlier scenario action, among other inconsistencies).

To address this issue, a scenario authoring system will be created. Authoring systems allows experts to design training situations to be played in the VTE [Dörner et al., 2015, Nagendran et al., 2015, Papelis et al., 2019]. The METIS authoring system will provide an interface to create classroom scenarios where students' behaviors are suggested to the expert creating the scenario. It will also automatically generate scenarios with the corresponding tags and will be able to prevent and detect inconsistencies (if a book is open earlier in the scenario, only the action to close that book will be suggested).

The authoring system will allow education experts to create their own scenarios and be able to adapt METIS for their own classroom. Before that, future work must determine the usability and UX of the authoring system in order to ensure that educators will be able to efficiently create new scenarios and use them in their classroom.

CHAPTER 6

CONCLUSION

The goal of this dissertation was to explore design principles for VTEs for teacher training. VTEs aim at providing training and learning through a virtual 3D environment. The implementation and study of VTEs are complex tasks as VTEs rely on a multitude of technical aspects such as computer graphics, visual display devices, interactions devices, and sound effects. These tasks become even more challenging when VTEs intend to train social skills. Our approach consisted of identifying guidelines for an efficient development of VTEs for social skills and to focus on the study of specific requirements including display and interaction devices and behavioral realism.

First, based on a review of the literature we established a set of six requirements for VTE for social skills training including behavioral fidelity, environment fidelity, instructional design, autonomy, interactivity, and scalability. For each requirement we suggested validation methods.

Second, we developed the IVT-T system over four development cycle using the established requirements. IVT-T is a virtual classroom simulator that plays scenarios validated by education experts with 3D virtual students. Teachers can use IVT-T to practice, reflect and receive feedback on their CBM techniques. The evaluation of IVT-T showed that IVT-T is a usable system offering realistic training situation supported by a solid instructional design. Additionally, IVT-T's virtual classrooms and students were evaluated as authentic by education experts.

Building on the final version of IVT-T we created METIS, a cross-platform simulator allowing to experience virtual classroom situations through immersive VR technologies (HMD and CAVE). METIS also enriched the set of behaviors of the student by enabling them to display non-verbal behaviors including facial expressions.

With METIS we proposed an approach to evaluate the usability and UX of cross-platforms systems. We observed that METIS was easy to use and to learn across all platforms. Additionally, our evaluation revealed that teacher students would quickly adopt METIS if it was used for their training.

Finally, using METIS, we explored the effect of VR platforms and display of facial expressions on users. Our results show that regardless of the display of facial expressions or of the platform, METIS provided an engaging experience in which participants were able to suspend their disbelief regarding the virtual students. During our study we also observed that facial expressions potentially generate a greater increase in the feeling of presence and co-presence when used with a PC platform. Based on these observations, we laid the ground of future research on the impact of facial expressions with immersive technologies.

APPENDIX A

QUESTIONNAIRES

QUIS

Questionnaire from [Chin et al., 1988].

OVERALL REACTION TO THE SOFTWARE		0	1	2	3	4	5	6	7	8	9		NA
1.	terrible											wonderful	
2.	difficult											easy	
3.	frustrating											satisfying	
4.	inadequate power											adequate power	
5.	dull											stimulating	
6.	rigid											flexible	
SCREEN		**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**		**NA**
7. Reading characters on the screen	hard											easy	
8. Highlighting simplifies task	not at all											very much	
9. Organization of information	confusing											very clear	
10. Sequence of screens	confusing											very clear	
TERMINOLOGY AND SYSTEM INFORMATION		**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**		**NA**
11. Use of terms throughout system	inconsistent											consistent	
12. Terminology related to task	never											always	
13. Position of messages on screen	inconsistent											consistent	
14. Prompts for input	confusing											clear	
15. Computer informs about its progress	never											always	
16. Error messages	unhelpful											helpful	
LEARNING		**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**		**NA**
17. Learning to operate the system	difficult											easy	
18. Exploring new features by trial and error	difficult											easy	
19. Remembering names and use of commands	difficult											easy	
20. Performing tasks is straightforward	never											always	
21. Help messages on the screen	unhelpful											helpful	
22. Supplemental reference materials	confusing											clear	
SYSTEM CAPABILITIES		**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**		**NA**
23. System speed	too slow											fast enough	
24. System reliability	unreliable											reliable	
25. System tends to be	noisy											quiet	
26. Correcting your mistakes	difficult											easy	
27. Designed for all levels of users	never											always	
		0	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**		**NA**

SUS

Questionnaire from [Brooke, 1986].

- I think that I would like to use this system frequently
- I found the system unnecessarily complex
- I thought the system was easy to use
- I think that I would need the support of a technical person to be able to use the system
- I found the various functions of the system were well integrated
- I thought there was too much inconsistency in the system
- I would imagine that most people would learn to use the system very quickly
- I found the system very awkward to use
- I felt very confident using the system
- I needed to learn a lot things before I could get going with the system

Technology Adoption

Questionnaire from [Tcha-Tokey et al., 2016].

- If I use again the same virtual environment, my interaction with the environment would be clear and understandable for me.
- It would be easy for me to become skillful at using the virtual environment.
- Learning to operate the virtual environment would be easy for me.
- Using the interaction devices (Virtual reality headset, CAVE and/or mouse) is a bad idea.
- The interaction devices (Virtual reality headset, CAVE and/or mouse) would make learning more interesting.

- I would like learning with the interaction devices (Virtual reality headset, CAVE and/or mouse).
- I have the resources necessary to use the interaction devices (Virtual reality headset, CAVE and/or mouse).

Experience Consequence

Questionnaire from [Tcha-Tokey et al., 2016].

- I suffered from fatigue during my interaction with the virtual environment.
- I suffered from headache during my interaction with the virtual environment.
- I suffered from eyestrain during my interaction with the virtual environment.
- I felt an increase of my salivation during my interaction with the virtual environment.
- I suffered from nausea during my interaction with the virtual environment.
- I suffered from fullness of the head during my interaction with the virtual environment.
- I suffered from dizziness with eye open during my interaction with the virtual environment.
- I suffered from vertigo during my interaction with the virtual environment.

User's Emotions

Questionnaire from [Tcha-Tokey et al., 2016].

- I enjoyed being in this virtual environment.
- It was so exciting that I could stay in the virtual environment for hours.
- I enjoyed the experience so much that I feel energized.

- I felt nervous in the virtual environment.
- I felt like distracting myself in order to reduce my anxiety.
- I found my mind wandering while I was in the virtual environment.
- The interaction devices (Virtual reality headset, CAVE and/or mouse) bored me to death.
- When my actions were going well, it gave me a rush.
- While using the interaction devices (Virtual reality headset, CAVE and/or mouse), I felt like time was dragging.
- I enjoyed the challenge of learning the virtual reality interaction devices ((Virtual reality headset, CAVE and/or mouse).
- I enjoyed dealing with the interaction devices (Virtual reality headset, CAVE and/or mouse).

Judgment

Questionnaire from [Tcha-Tokey et al., 2016].

- I found that this virtual environment was Lame/Exciting
- I found this virtual environment Amateurish/Professional
- I found this virtual environment Gaudy/Classy
- I found that this virtual environment is Ugly/Beautiful
- I found that this virtual environment is Disagreeable/Likable

Presence

Questionnaire from [Bailenson et al., 2005].

- I forgot about my immediate physical surroundings (i.e., the lab room) when I was in the virtual classroom.
- I paid more attention to my own thoughts (e.g., personal preoccupations, daydreams, etc.) than what was going on in the virtual classroom.
- I did not want to reach out and touch things in the virtual classroom.
- I felt like I was in a psychology laboratory rather than a virtual classroom.

Co-Presence

Questionnaire from [Bailenson and Yee, 2006].

- I perceived that I was in the presence of students in the virtual classroom with me.
- I felt that the students in the virtual classroom were watching me and were aware of my presence.
- The thought that they were not real students crossed my mind often in the virtual classroom.
- The students in the virtual classroom appeared to be sentient (conscious and alive) to me.
- I perceived the students as being only a computerized image, not as real students.

Engagement

Questionnaire from [O'Brien et al., 2018].

- I lost myself in this experience.
- The time I spent using the classroom simulator just slipped away.

- I was absorbed in this experience.
- I felt frustrated while using this classroom simulator.
- I found this classroom simulator confusing to use.
- Using this classroom simulator was taxing.
- This classroom simulator was attractive.
- This classroom simulator was aesthetically appealing.
- This classroom simulator appealed to my senses.
- Using classroom simulator was worthwhile.
- My experience was rewarding.
- I felt interested in this experience.

Believability

Questionnaire adapted from [Gomes et al., 2013].

- The disruptive student perceives the world around him/her.
- It is easy to understand what the disruptive student is thinking about.
- The disruptive student has a personality.
- The disruptive student's behavior draws my attention.
- The disruptive student's behavior is predictable.
- The disruptive student's behavior is coherent.
- The disruptive student's behavior changes according to experience.
- The disruptive student interacts socially with other characters.

BIBLIOGRAPHY

[Amini and Lisetti, 2013] Amini, R. and Lisetti, C. (2013). Hapfacs: An open source api/software to generate facs-based expressions for ecas animation and for corpus generation. In *2013 Humaine Association Conference on Affective Computing and Intelligent Interaction*, pages 270–275. IEEE.

[Andrade and Law, 2018] Andrade, P. and Law, E. L.-C. (2018). User-based evaluation of gamification elements in an educational application. In *Proceedings of the 32nd International BCS Human Computer Interaction Conference 32*, pages 1–13.

[Badiee and Kaufman, 2015] Badiee, F. and Kaufman, D. (2015). Design evaluation of a simulation for teacher education. *Sage Open*, 5(2).

[Bailenson et al., 2005] Bailenson, J. N., Swinth, K., Hoyt, C., Persky, S., Dimov, A., and 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.

[Bailenson and Yee, 2006] Bailenson, J. N. and Yee, N. (2006). A longitudinal study of task performance, head movements, subjective report, simulator sickness, and transformed social interaction in collaborative virtual environments. *Presence: Teleoperators and Virtual Environments*, 15(6):699–716.

[Bangor et al., 2008] Bangor, A., Kortum, P. T., and Miller, J. T. (2008). An empirical evaluation of the system usability scale. *Intl. Journal of Human–Computer Interaction*, 24(6):574–594.

[Barmaki and Hughes, 2015] Barmaki, R. and Hughes, C. E. (2015). Providing real-time feedback for student teachers in a virtual rehearsal environment. In *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction*, pages 531–537.

[Baron, 1987] Baron, R. A. (1987). Interviewer's moods and reactions to job applicants: The influence of affective states on applied social judgments 1. *Journal of Applied Social Psychology*, 17(10):911–926.

[Barot et al., 2013] Barot, C., Lourdeaux, D., Burkhardt, J.-M., Amokrane, K., and Lenne, D. (2013). V3s: A virtual environment for risk-management training based on human-activity models. *Presence*, 22(1):1–19.

[Bartneck, 2003] Bartneck, C. (2003). Interacting with an embodied emotional character. In *Proceedings of the 2003 international conference on Designing pleasurable products and interfaces*, pages 55–60.

[Bates et al., 1994] Bates, J. et al. (1994). The role of emotion in believable agents. *Communications of the ACM*, 37(7):122–125.

[Baylor et al., 2003] Baylor, A., Shen, E., and Huang, X. (2003). Which pedagogical agent do learners choose? the effects of gender and ethnicity. In *E-Learn: World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education*, pages 1507–1510. Association for the Advancement of Computing in Education (AACE).

[Beale and Creed, 2009] Beale, R. and Creed, C. (2009). Affective interaction: How emotional agents affect users. *International journal of human-computer studies*, 67(9):755–776.

[Becker-Asano, 2014] Becker-Asano, C. (2014). Wasabi for affect simulation in human-computer interaction. In *Proc. International Workshop on Emotion Representations and Modelling for HCI Systems*. Citeseer.

[Berry et al., 2005] Berry, D. C., Butler, L. T., and De Rosis, F. (2005). Evaluating a realistic agent in an advice-giving task. *International Journal of Human-Computer Studies*, 63(3):304–327.

[Bevacqua et al., 2007] Bevacqua, E., Mancini, M., Niewiadomski, R., and Pelachaud, C. (2007). An expressive eca showing complex emotions. In *Proceedings of the AISB annual convention, Newcastle, UK*, pages 208–216.

[Billinghurst and Weghorst, 1995] Billinghurst, M. and Weghorst, S. (1995). The use of sketch maps to measure cognitive maps of virtual environments. In *Proceedings Virtual Reality Annual International Symposium'95*, pages 40–47. IEEE.

[Birns et al., 2002] Birns, J. H., Joffre, K. A., Leclerc, J. F., and Paulsen, C. A. (2002). Getting the whole picture: Collecting usability data using two methods—- concurrent think aloud and retrospective probing. In *Proceedings of UPA Conference*, pages 8–12. Citeseer.

[Blom and Beckhaus, 2014] Blom, K. J. and Beckhaus, S. (2014). The design space of dynamic interactive virtual environments. *Virtual Reality*, 18(2):101–116.

[Bogdanovych et al., 2016] Bogdanovych, A., Trescak, T., and Simoff, S. (2016). What makes virtual agents believable? *Connection Science*, 28(1):83–108.

[Bossard et al., 2008] Bossard, C., Kermarrec, G., Buche, C., and Tisseau, J. (2008). Transfer of training in virtual environments: A new challenge. *Virtual Reality*, (12):151–161.

[Bowman, 1998] Bowman, D. A. (1998). Interaction techniques for immersive virtual environments: Design, evaluation, and application. *Journal of Visual Languages and Computing*, 10:37–53.

[Bowman et al., 2002a] Bowman, D. A., Datey, A., Ryu, Y. S., Farooq, U., and Vasnaik, O. (2002a). Empirical comparison of human behavior and performance with different display devices for virtual environments. In *Proceedings of the human factors and ergonomics society annual meeting*, volume 46, pages 2134–2138. SAGE Publications Sage CA: Los Angeles, CA.

[Bowman et al., 2002b] Bowman, D. A., Gabbard, J. L., and Hix, D. (2002b). A survey of usability evaluation in virtual environments: classification and comparison of methods. *Presence: Teleoperators & Virtual Environments*, 11(4):404–424.

[Brenton et al., 2005] Brenton, H., Gillies, M., Ballin, D., and Chatting, D. (2005). The uncanny valley: does it exist. In *Proceedings of conference of human computer interaction, workshop on human animated character interaction*. Citeseer.

[Brooke, 1986] Brooke, J. (1986). System usability scale (sus): a quick-and-dirty method of system evaluation user information. *Reading, UK: Digital Equipment Co Ltd*, page 43.

[Buttussi and Chittaro, 2018] Buttussi, F. and Chittaro, L. (2018). Effects of different types of virtual reality display on presence and learning in a safety training scenario. *IEEE transactions on visualization and computer graphics*, 24(2):1063–1076.

[Cabral et al., 2017] Cabral, J. P., Cowan, B. R., Zibrek, K., and McDonnell, R. (2017). The influence of synthetic voice on the evaluation of a virtual character. In *INTERSPEECH*, pages 229–233.

[Cao et al., 2019] Cao, L., Peng, C., and Hansberger, J. T. (2019). Usability and engagement study for a serious virtual reality game of lunar exploration missions. In *Informatics*, volume 6, page 44. Multidisciplinary Digital Publishing Institute.

[Casanueva and Blake, 2001] Casanueva, J. and Blake, E. (2001). The effects of avatars on co-presence in a collaborative virtual environment. Paper presented at the South African Institute of Computer Scientists and Information Technologists Annual Conference, Pretoria, South Africa.

[Chellali et al., 2016] Chellali, A., Mentis, H., Miller, A., Ahn, W., Arikatla, V. S., Sankaranarayanan, G., De, S., Schwaitzberg, S. D., and 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.

[Chen and Bowman, 2009] Chen, J. and Bowman, D. A. (2009). Domain-specific design of 3d interaction techniques: An approach for designing useful virtual environment applications. *Presence: Teleoperators and Virtual Environments*, 18(5):370–386.

[Chin et al., 1988] Chin, J. P., Diehl, V. A., and Norman, K. L. (1988). Development of an instrument measuring user satisfaction of the human-computer interface. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 213–218. ACM.

[Christensen et al., 2011] Christensen, R., Knezek, G., Wood, T. T., Gibson, D., Tyler-Wood, T., Gibson, D., Knezek, G., and Gibson, D. (2011). SimSchool: An online dynamic simulator for enhancing teacher preparation. *International Journal of Learning Technology*.

[Christianson and Loftus, 1991] Christianson, S.-Å. and Loftus, E. F. (1991). Remembering emotional events: The fate of detailed information. *Cognition & Emotion*, 5(2):81–108.

[Christou and Aristidou, 2017] Christou, C. G. and Aristidou, P. (2017). Steering versus teleport locomotion for head mounted displays. In *International Conference on Augmented Reality, Virtual Reality and Computer Graphics*, pages 431–446. Springer.

[Cialdini, 2009] Cialdini, R. B. (2009). *Influence: Science and practice*, volume 4. Pearson education Boston.

[Collum et al., 2020] Collum, D., Christensen, R., Delicath, T., and Knezek, G. (2020). Measuring changes in educator bias in a simulated learning environment. In *Society for Information Technology & Teacher Education International Conference*, pages 496–502. Association for the Advancement of Computing in Education (AACE).

[Cooke, 2010] 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.

[Corradini, 2004] Corradini, A. (2004). Collaborative integration of speech and 3d gesture for map-based applications. In *International Conference on Computational Science*, pages 913–920. Springer.

[Cowell and Stanney, 2005] Cowell, A. J. and Stanney, K. M. (2005). Manipulation of non-verbal interaction style and demographic embodiment to increase anthropomorphic computer character credibility. *International journal of human-computer studies*, 62(2):281–306.

[Creemers, 1994] Creemers, B. P. M. (1994). Effective instruction as a basis for effective education in schools. *Tijdschrift Voor Onderwijsresearch*, 19(1):3–16.

[Creswell and Plano Clark, 2011] Creswell, J. W. and Plano Clark, V. (2011). Choosing a mixed methods design. *Designing and conducting mixed methods research*, 2:53–106.

[Crone and Teddlie, 1995] Crone, L. J. and Teddlie, C. (1995). Further examination of teacher behavior in differentially effective schools: Selection and socialization process. *Journal of Classroom Interaction*, 30(1):1–9.

[Cummings and Bailenson, 2016] Cummings, J. J. and Bailenson, J. N. (2016). How immersive is enough? a meta-analysis of the effect of immersive technology on user presence. *Media Psychology*, 19(2):272–309.

[Dalgarno et al., 2016] 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 and Lee, 2009] Dalgarno, B. and Lee, M. J. (2009). What are the learning affordances of 3-D virtual environments? *British Journal of Educational Technology*, 41(1):10–32.

[Dautenhahn, 1998] Dautenhahn, K. (1998). The art of designing socially intelligent agents: Science, fiction, and the human in the loop. *Applied artificial intelligence*, 12(7-8):573–617.

[Davis, 1993] Davis, A. M. (1993). *Software requirements: objects, functions, and states.* Prentice-Hall, Inc.

[Dawson and Lignugaris/Kraft, 2017] Dawson, M. R. and Lignugaris/Kraft, B. (2017). Meaningful practice: Generalizing foundation teaching skills from tle teachlive^{TM} to the classroom. *Teacher Education and Special Education*, 40(1):26–50.

[De Freitas and Oliver, 2006] De Freitas, S. and 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 et al., 2010] De Freitas, S., Rebolledo-Mendez, G., Liarokapis, F., Magoulas, G., and Poulovassilis, 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 et al., 2017] Debattista, K., Bashford-Rogers, T., Harvey, C., Waterfield, B., and Chalmers, A. (2017). Subjective evaluation of high-fidelity virtual environments for driving simulations. *IEEE Transactions on Human-Machine Systems*, 48(1):30–40.

[Delamarre et al., 2019a] Delamarre, A., Buche, C., and Lisetti, C. (2019a). Aimer: Appraisal interpersonal model of emotion regulation, affective virtual students to support teachers training. In *Proceedings of the 19th ACM International Conference on Intelligent Virtual Agents*, pages 182–184.

[Delamarre et al., 2020a] Delamarre, A., Buche, C., and Lisetti, C. (2020a). Modeling emotions for training in immersive simulations (metis): a cross-platform virtual classroom study. In *2020 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*, pages 78–84.

[Delamarre et al., 2017] Delamarre, A., Buche, C., Polceanu, M., Lunn, S., Ruiz, G., Bolivar, S., Shernoff, E., and Lisetti, C. (2017). An Interactive Virtual Training (IVT) simulation for early career teachers to practice in 3D classrooms with student avatars. In *FLAIRS 2017 - Proceedings of the 30th International Florida Artificial Intelligence Research Society Conference*.

[Delamarre et al., 2020b] Delamarre, A., Lisetti, C., and Buche, C. (2020b). A cross-platform classroom training simulator: Interaction design and evaluation. In *2020 International Conference on Cyberworlds (CW)*, pages 86–93. IEEE.

[Delamarre et al., 2019b] Delamarre, A., Lunn, S., Buche, C., Shernoff, E., Frazier, S., and Lisetti, C. (2019b). Interdisciplinary collaboration and establishment of requirements for a 3d interactive virtual training for teachers. In *Proceedings of the 19th ACM International Conference on Intelligent Virtual Agents*, pages 185–187.

[Denton and Hasbrouck, 2009] Denton, C. A. and Hasbrouck, J. (2009). A description of instructional coaching and its relationship to consultation. *Journal of Educational and Psychological Consultation*, 19:150–175.

[Dias et al., 2014] Dias, J., Mascarenhas, S., and Paiva, A. (2014). Fatima modular: Towards an agent architecture with a generic appraisal framework. In *Emotion modeling*, pages 44–56. Springer.

[Dieker et al., 2015] Dieker, L. A., Hynes, M. C., Hughes, C. E., Hardin, S., and Becht, K. (2015). TLE TeachLivE™: Using Technology to Provide Quality Professional Development in Rural Schools. *Rural Special Education Quarterly*.

[Dörner et al., 2015] Dörner, R., Kallmann, M., and Huang, Y. (2015). Content creation and authoring challenges for virtual environments: from user interfaces to autonomous virtual characters. In *Virtual Realities*, pages 187–212. Springer.

[Drettakis et al., 2007] Drettakis, G., Roussou, M., Reche, A., and 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.

[Ekman, 2004] Ekman, P. (2004). Emotions revealed: recognizing faces and feelings to improve communication and emotional life-henry holt & company inc. *New York, United States*.

[Ekman, 1997] Ekman, R. (1997). *What the face reveals: Basic and applied studies of spontaneous expression using the Facial Action Coding System (FACS)*. Oxford University Press, USA.

[Evans, 2002] Evans, D. (2002). *Emotion: The science of sentiment*. Oxford University Press, USA.

[Evertson and Weinstein, 2013] Evertson, C. M. and Weinstein, C. S. (2013). *Handbook of classroom management: Research, practice, and contemporary issues*. Routledge.

[Fabri et al., 2005] Fabri, M., Moore, D., and Hobbs, D. (2005). Empathy and enjoyment in instant messaging. In *Proceedings of 19th British HCI group annual conference (HCI2005), Edinburgh, UK*, pages 4–9.

[Felnhofer et al., 2019] Felnhofer, A., Kaufmann, M., Atteneder, K., Kafka, J. X., Hlavacs, H., Beutl, L., Hennig-Fast, K., and Kothgassner, O. D. (2019). The mere presence of an attentive and emotionally responsive virtual character influences focus of attention and perceived stress. *International Journal of Human-Computer Studies*, 132:45–51.

[Fisher et al., 2011] Fisher, M. J., Marshall, A. P., and Mitchell, M. (2011). Testing differences in proportions. *Australian Critical Care*, 24(2):133–138.

[Gao et al., 2015] Gao, Z., Yu, Y., Zhou, Y., and Du, S. (2015). Leveraging two kinect sensors for accurate full-body motion capture. *Sensors*, 15(9):24297–24317.

[Gerbaud et al., 2008] Gerbaud, S., Mollet, N., Ganier, F., Arnaldi, B., and Tisseau, J. (2008). GVT: A platform to create virtual environments for procedural training. In *Proceedings - IEEE Virtual Reality*, pages 225–232.

[Gibson, 2011] Gibson, D. (2011). Modeling emotions in simulated learning. *Standards in Emotion Modeling*.

[Gomes et al., 2013] Gomes, P., Paiva, A., Martinho, C., and Jhala, A. (2013). Metrics for character believability in interactive narrative. In *International Conference on Interactive Digital Storytelling*, pages 223–228. Springer.

[Gratch and Marsella, 2004] Gratch, J. and Marsella, S. (2004). A domain-independent framework for modeling emotion. *Cognitive Systems Research*, 5(4):269–306.

[Gregory et al., 2011] Gregory, S., Dalgarno, B., Campbell, M., Reiners, T., Knox, V., and Masters, Y. (2011). Changing directions through VirtualPREX : engaging pre-service teachers in virtual professional experience. In *ASCILITE*, pages 491–501.

[Gregory et al., 2013] Gregory, S., Dalgarno, B., Crisp, G., Reiners, T., Masters, Y., Dreher, H., and Knox, V. (2013). *VirtualPREX: Innovative assessment using a 3D virtual world with pre-service teachers*. Office for Learning and Teaching.

[Gross, 2015] Gross, J. J. (2015). The extended process model of emotion regulation: Elaborations, applications, and future directions. *Psychological Inquiry*, 26(1):130–137.

[Grossman and McDonald, 2008] Grossman, P. and McDonald, M. (2008). Back to the future: Directions for research in teaching and teacher education. *American Educational Research Journal*, 45:184–205.

[Gupta et al., 2008] Gupta, S., Anand, D., Brough, J., Schwartz, M., and Kavetsky, R. (2008). Training in virtual environments: A safe, cost effective, and engaging approach to training.

[Hamdy and King, 2017] Hamdy, S. and King, D. (2017). Affect and believability in game characters – a review of the use of affective computing in games. pages 90–97.

[Harjunen et al., 2018] Harjunen, V. J., Spapé, M., Ahmed, I., Jacucci, G., and Ravaja, N. (2018). Persuaded by the machine: The effect of virtual nonverbal cues and individual differences on compliance in economic bargaining. *Computers in Human Behavior*, 87:384–394.

[Hartholt et al., 2019] Hartholt, A., Mozgai, S., Fast, E., Liewer, M., Reilly, A., Whitcup, W., and Rizzo, A. S. (2019). Virtual humans in augmented reality: A first step towards real-world embedded virtual roleplayers. In *Proceedings of the 7th International Conference on Human-Agent Interaction*, pages 205–207.

[Hartson and Pyla, 2013] Hartson, R. and Pyla, P. A. (2013). The UX book: Process and guidelines for ensuring a quality user experience. *Communication Design Quarterly Review*, 2(1):67–72.

[Hartson and Pyla, 2018] Hartson, R. and Pyla, P. S. (2018). *The UX book: Agile UX design for a quality user experience*. Morgan Kaufmann.

[Hassenzahl et al., 2003] Hassenzahl, M., Burmester, M., and Koller, F. (2003). Attrakdiff: A questionnaire to measure perceived hedonic and pragmatic quality. In *Mensch & Computer*, volume 57, pages 187–196.

[Hayes et al., 2013a] Hayes, A. T., Hardin, S. E., and Hughes, C. E. (2013a). Perceived presence's role on learning outcomes in a mixed reality classroom of simulated students. In *International Conference on Virtual, Augmented and Mixed Reality*, pages 142–151. Springer.

[Hayes et al., 2013b] Hayes, A. T., Straub, C. L., Dieker, L. A., Hughes, C. E., and Hynes, M. C. (2013b). Ludic learning: Exploration of tle teachlive^{TM} and effective teacher training. *International Journal of Gaming and Computer-Mediated Simulations (IJGCMS)*, 5(2):20–33.

[Heeter, 1992] Heeter, C. (1992). Being there: The subjective experience of presence. *Presence: Teleoperators & Virtual Environments*, 1(2):262–271.

[Hendrix and Barfield, 1996] Hendrix, C. and Barfield, W. (1996). Presence within virtual environments as a function of visual display parameters. *Presence: Teleoperators & Virtual Environments*, 5(3):274–289.

[Henry et al., 2011a] Henry, D. B., Farrell, A. D., Schoeny, M. E., Tolan, P. H., and Dymnicki, A. B. (2011a). Influence of school-level variables on aggression and associated attitudes of middle school students. *Journal of School Psychology*, 49(5):481–503.

[Henry et al., 2011b] Henry, D. B., Farrell, A. D., Schoeny, M. E., Tolan, P. H., and Dymnicki, A. B. (2011b). Influence of school-level variables on aggression and associated attitudes of middle school students. *Journal of school psychology*, 49(5):481–503.

[Hirose et al., 2009] Hirose, M., Schmalstieg, D., Wingrave, C., and Nishimura, K. (2009). Higher levels of immersion improve procedure memorization performance. In *Proceedings of the 15th Joint virtual reality Eurographics conference on Virtual Environments*, pages 121–128.

[IJsselsteijn et al., 2013] IJsselsteijn, W., de Kort, Y., and Poels, K. (2013). *The Game Experience Questionnaire*.

[Ingersoll and Smith, 2003] Ingersoll, R. M. and Smith, T. M. (2003). The wrong solution to the teacher shortage. *Educational Leadership*, 60(8):30–33.

[iPi Soft LLC, 2020] iPi Soft LLC (2008-2020). ipisoft, motion capture for the masses.

[Isen et al., 1987] Isen, A. M., Daubman, K. A., and Nowicki, G. P. (1987). Positive affect facilitates creative problem solving. *Journal of personality and social psychology*, 52(6):1122.

[Jaspers, 2009] 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 et al., 2018] Jemmali, C., Bunian, S., Mambretti, A., and El-Nasr, M. S. (2018). Educational game design: an empirical study of the effects of narrative. In

Proceedings of the 13th International Conference on the Foundations of Digital Games, page 34. ACM.

[Johnsen and Lok, 2008] Johnsen, K. and Lok, B. (2008). An evaluation of immersive displays for virtual human experiences. In *Virtual Reality Conference, 2008. VR'08. IEEE*, pages 133–136. IEEE.

[Juan and Pérez, 2009] Juan, M. C. and Pérez, D. (2009). Comparison of the levels of presence and anxiety in an acrophobic environment viewed via hmd or cave. *Presence: Teleoperators and Virtual Environments*, 18(3):232–248.

[Junod et al., 2006] Junod, R. E. V., DuPaul, G. J., Jitendra, A. K., Volpe, R. J., and 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.

[Kazdin, 2008] Kazdin, A. E. (2008). *Parent management training: Treatment for oppositional, aggressive, and antisocial behavior in children and adolescents*. Oxford University Press.

[Kelleci and Aksoy, 2020] Kelleci, Ö. and Aksoy, N. C. (2020). Using game-based virtual classroom simulation in teacher training: User experience research. *Simulation & Gaming*.

[Keller and Suzuki, 2004] Keller, J. and Suzuki, K. (2004). Learner motivation and e-learning design: A multinationally validated process. *Journal of educational Media*, 29(3):229–239.

[Kennedy et al., 1993] Kennedy, R. S., Lane, N. E., Berbaum, K. S., and Lilienthal, M. G. (1993). Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness. *The international journal of aviation psychology*, 3(3):203–220.

[Kervin et al., 2006] Kervin, L., Ferry, B., and Carrington, L. (2006). Classsim: preparing tomorrows teachers for classroom reality. In *Society for Information Technology & Teacher Education International Conference*, pages 3204–3211. Association for the Advancement of Computing in Education (AACE).

[Kim et al., 2014] Kim, K., Rosenthal, M. Z., Zielinski, D. J., and Brady, R. (2014). Effects of virtual environment platforms on emotional responses. *Computer methods and programs in biomedicine*, 113(3):882–893.

[Köknar, 2015] Köknar, C. (2015). Modeling student behaviours in a virtual classroom with incorporation of social learning theory into belief-desire-intention model (yaymlanmam yüksek lisans tezi). *Orta Dou Teknik Üniversitesi/GATE, Ankara.*

[Kopp et al., 2006] Kopp, S., Krenn, B., Marsella, S., Marshall, A. N., Pelachaud, C., Pirker, H., Thórisson, K. R., and Vilhjálmsson, H. (2006). Towards a common framework for multimodal generation: The behavior markup language. In *International workshop on intelligent virtual agents*, pages 205–217. Springer.

[Kwon et al., 2013] Kwon, J. H., Powell, J., and Chalmers, A. (2013). How level of realism influences anxiety in virtual reality environments for a job interview. *International Journal of Human Computer Studies*.

[Lazarus and Lazarus, 1991] Lazarus, R. S. and Lazarus, R. S. (1991). *Emotion and adaptation*. Oxford University Press on Demand.

[Lee, 2004] Lee, K. M. (2004). Presence, explicated. *Communication theory*, 14(1):27–50.

[Lee and Heeter, 2015] Lee, M. S. and Heeter, C. (2015). Cognitive intervention and reconciliation: Npc believability in single-player rpgs. *International Journal of Role-Playing*, (5):47–65.

[Lester et al., 2000] Lester, J., Towns, S., and Callaway, C. (2000). Deictic and emotive communication in animated pedagogical agents. *Embodied Conversational Agents*, pages 123–154.

[Leventhal, 1984] Leventhal, H. (1984). A perceptual-motor theory of emotion. In *Advances in experimental social psychology*, volume 17, pages 117–182. Elsevier.

[Liew et al., 2017] Liew, T. W., Zin, N. A. M., and Sahari, N. (2017). Exploring the affective, motivational and cognitive effects of pedagogical agent enthusiasm in a multimedia learning environment. *Human-centric Computing and Information Sciences*, 7(1):9.

[Lisetti et al., 2013] Lisetti, C., Amini, R., Yasavur, U., and Rishe, N. (2013). I can help you change! an empathic virtual agent delivers behavior change health interventions. *ACM Transactions on Management Information Systems*, 4(4):1–28.

[Llorach et al., 2019] Llorach, G., Agenjo, J., Blat, J., and Sayago, S. (2019). Web-based embodied conversational agents and older people. In *Perspectives on Human-Computer Interaction Research with Older People*, pages 119–135. Springer.

[Loyall, 1997] Loyall, A. B. (1997). Believable agents: Building interactive personalities. Technical report, CARNEGIE-MELLON UNIV PITTSBURGH PA DEPT OF COMPUTER SCIENCE.

[LucidChart, 2020] LucidChart (2020). Lucid chart.

[Ludwick and Doucette, 2009] Ludwick, D. A. and 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 et al., 2019] Lugrin, J.-L., Bosser, A.-G., Latoschik, M. E., Chollet, M., Glémarec, Y., and Lugrin, B. (2019). Towards narrative-driven atmosphere for virtual classrooms. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*, pages 1–6.

[Lugrin et al., 2013] Lugrin, J.-L., Cavazza, M., Charles, F., Le Renard, M., Freeman, J., and Lessiter, J. (2013). Immersive fps games: user experience and performance. In *Proceedings of the 2013 ACM international workshop on Immersive media experiences*, pages 7–12.

[Lugrin et al., 2016] Lugrin, J.-L., Latoschik, M. E., Habel, M., Roth, D., Seufert, C., and Grafe, S. (2016). Breaking Bad Behaviors: A New Tool for Learning Classroom Management Using Virtual Reality. *Frontiers in ICT*.

[Magnentat-Thalman and Thalmann, 2005] Magnentat-Thalman, N. and Thalmann, D. (2005). Virtual humans: thirty years of research, what next? *The Visual Computer*, 21(12):997–1015.

[MakeHuman, 2014] MakeHuman (2014). Make human, open source tool for making 3d characters.

[Malatesta et al., 2009] Malatesta, L., Raouzaiou, A., Karpouzis, K., and Kollias, S. (2009). Mpeg-4 facial expression synthesis. *Personal and ubiquitous computing*, 13(1):77–83.

[Malone, 1980] Malone, T. W. (1980). What makes things fun to learn? heuristics for designing instructional computer games. In *Proceedings of the 3rd ACM SIGSMALL symposium and the first SIGPC symposium on Small systems*, pages 162–169. ACM.

[Mateas, 1999] Mateas, M. (1999). An oz-centric review of interactive drama and believable agents. In *Artificial intelligence today*, pages 297–328. Springer.

[McLaughlin et al., 2010] McLaughlin, T., Smith, D., and Brown, I. A. (2010). A framework for evidence based visual style development for serious games. In *Proceedings of the Fifth International Conference on the Foundations of Digital Games - FDG '10*, pages 132–138.

[McMahan et al., 2012] McMahan, R. P., Bowman, D. A., Zielinski, D. J., and Brady, R. B. (2012). Evaluating display fidelity and interaction fidelity in a virtual reality game. *IEEE transactions on visualization and computer graphics*, 18(4):626–633.

[Merrill, 2009] Merrill, M. D. (2009). Instructional-design theories and models: Building a common knowledge base. *A Reigeluth CM & Carr-Chellman (ed.). Lawrence Erlbaum Associates*, pages 41–56.

[Mikropoulos and Natsis, 2011] Mikropoulos, T. A. and Natsis, A. (2011). Educational virtual environments: A ten-year review of empirical research (1999–2009). *Computers & Education*, 56(3):769–780.

[Millais et al., 2018] Millais, P., Jones, S. L., and Kelly, R. (2018). Exploring data in virtual reality: Comparisons with 2d data visualizations. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*, pages 1–6.

[Mizell et al., 2002] Mizell, D. W., Jones, S. P., Slater, M., and Spanlang, B. (2002). Comparing immersive virtual reality with other display modes for visualizing complex 3d geometry. *Boeing and University College*.

[Moreno and Flowerday, 2006] Moreno, R. and Flowerday, T. (2006). Students' choice of animated pedagogical agents in science learning: A test of the similarity-attraction hypothesis on gender and ethnicity. *Contemporary educational psychology*, 31(2):186–207.

[Morgan, 1996] Morgan, D. L. (1996). *Focus groups as qualitative research*, volume 16. Sage publications.

[Moridis and Economides, 2012] Moridis, C. N. and Economides, A. A. (2012). Affective learning: Empathetic agents with emotional facial and tone of voice expressions. *IEEE Transactions on Affective Computing*, 3(3):260–272.

[Moskaliuk et al., 2013] Moskaliuk, J., Bertram, J., and Cress, U. (2013). Training in virtual environments: putting theory into practice. *Ergonomics*, 56(2):195–204.

[Mount et al., 2009] Mount, N., Chambers, C., Weaver, D., and Priestnall, G. (2009). Learner immersion engagement in the 3d virtual world: principles emerging from the delve project. *Innovation in Teaching and Learning in Information and Computer Sciences*, 8(3):40–55.

[Nadolski et al., 2012] Nadolski, R. J., Hummel, H. G., Slootmaker, A., and Van Der Vegt, W. (2012). Architectures for developing multiuser, immersive learning scenarios. *Simulation & Gaming*, 43(6):825–852.

[Nagendran et al., 2014] Nagendran, A., Pillat, R., Kavanaugh, A., Welch, G., and Hughes, C. (2014). A unified framework for individualized avatar-based interactions. *Presence: Teleoperators and Virtual Environments*, 23(2):109–132.

[Nagendran et al., 2015] Nagendran, A., Welch, G., Hughes, C., and Pillat, R. (2015). Technical report: exploring human surrogate characteristics. In *Virtual Realities*, pages 215–228. Springer.

[Nakayama et al., 2015] Nakayama, S., Jin, G., et al. (2015). Safety training enhancing outcomes through virtual environments. *Professional Safety*, 60(02):34–38.

[Nakhal et al., 2016] Nakhal, B., Bevacqua, E., Said, B., and Querrec, R. (2016). Cognitive architecture for embodied conversational agent: Application on virtual learning environment. pages 4973–4983.

[Napa et al., 2019] Napa, S., Moore, M., and Bardyn, T. (2019). Advancing cardiac surgery case planning and case review conferences using virtual reality in medical libraries: evaluation of the usability of two virtual reality apps. *JMIR human factors*, 6(1):e12008.

[Nielsen, 1994] Nielsen, J. (1994). *Usability engineering*. Elsevier.

[Niven et al., 2009] Niven, K., Totterdell, P., and Holman, D. (2009). A classification of controlled interpersonal affect regulation strategies. *Emotion*, 9(4):498.

[O'Brien and Toms, 2008] O'Brien, H. L. and Toms, E. G. (2008). What is user engagement? a conceptual framework for defining user engagement with technology. *Journal of the American society for Information Science and Technology*, 59(6):938–955.

[Ochs et al., 2018] Ochs, M., Mestre, D., De Montcheuil, G., Pergandi, J.-M., Saubesty, J., Lombardo, E., Francon, D., and 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*, pages 1–11.

[Oliver and Reschly, 2007] Oliver, R. M. and Reschly, D. J. (2007). *Effective classroom management: Teacher preparation and professional development.* Washington, DC: National Comprehensive Center for Teacher Quality.

[Oliver and Reschly, 2010] Oliver, R. M. and 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.

[Ortony, 2003] Ortony, A. (2003). On making believable emotional agents believable. trapple, rp ed.: Emotions in humans and artefacts.

[Ortony et al., 1988] Ortony, A., Clore, G., and Collins, A. (1988). *The Cognitive Structure of Emotion*, volume 18.

[O'Brien et al., 2018] O'Brien, H. L., Cairns, P., and Hall, M. (2018). A practical approach to measuring user engagement with the refined user engagement scale (ues) and new ues short form. *International Journal of Human-Computer Studies*, 112:28–39.

[Pan et al., 2011] Pan, X., Banakou, D., and Slater, M. (2011). Computer based video and virtual environments in the study of the role of emotions in moral behavior. In *Affective computing and intelligent interaction*, pages 52–61. Springer.

[Pandey and Pandey, 2012] Pandey, D. and Pandey, V. (2012). Importance of requirement management : A requirement engineering concern. *International Journal of Research and Development - A Management Review (IJRDMR)*, 1:2319–5479.

[Papelis et al., 2019] Papelis, Y., Watson, G., and Croll, M. (2019). Framework for authoring repeatable scenarios in virtual environments populated with stochastic virtual agents. In *Proceedings of the 19th ACM International Conference on Intelligent Virtual Agents*, pages 221–223.

[Pawel et al., 2009] Pawel, D., Michal, P., Rafal, R., and Kenji, A. (2009). Humoroids: conversational agents that induce positive emotions with humor. In *AAMAS'09 Proceedings of The 8th International Conference on Autonomous Agents and Multiagent Systems*, volume 2, pages 1171–1172. ACM.

[Pekrun et al., 2011] Pekrun, R., Goetz, T., Frenzel, A. C., Barchfeld, P., and Perry, R. P. (2011). Measuring emotions in students' learning and performance: The achievement emotions questionnaire (aeq). *Contemporary educational psychology*, 36(1):36–48.

[Perneger, 1998] Perneger, T. V. (1998). What's wrong with bonferroni adjustments. *Bmj*, 316(7139):1236–1238.

[Poel et al., 2009] Poel, M., Heylen, D., Nijholt, A., Meulemans, M., and Van Breemen, A. (2009). Gaze behaviour, believability, likability and the icat. *AI & society*, 24(1):61–73.

[Querrec et al., 2004] Querrec, R., Buche, C., Maffre, E., and Chevaillier, P. (2004). Multiagents systems for virtual environment for training. application to fire-fighting. *International Journal of Computers and Applications*, 1(1):25–34.

[Ragan et al., 2015] Ragan, E., Bowman, D., Kopper, R., Stinson, C., Scerbo, S., and McMahan, R. (2015). Effects of field of view and visual complexity on virtual reality training effectiveness for a visual scanning task. *IEEE transactions on visualization and computer graphics*, 21(7):794–807.

[Ragan et al., 2012] Ragan, E. D., Wood, A., McMahan, R. P., and Bowman, D. A. (2012). Trade-offs related to travel techniques and level of display fidelity in virtual data-analysis environments. In *ICAT/EGVE/EuroVR*, pages 81–84.

[Ravenet et al., 2016] Ravenet, B., Bevacqua, E., Cafaro, A., Ochs, M., and Pelachaud, C. (2016). Perceiving attitudes expressed through nonverbal behaviors in immersive virtual environments. In *Proceedings of the 9th International Conference on Motion in Games*, pages 175–180. ACM.

[Rayner and Fluck, 2014] Rayner, C. and 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 et al., 2011] Richey, R., Klein, J., and Tracey, M. (2011). Chapter 2: General systems theory. *The Instructional Design Knowledge Base: Theory, Research and Practice*, pages 11–29.

[Robb et al., 2015] Robb, A., White, C., Cordar, A., Wendling, A., Lampotang, S., and Lok, B. (2015). A comparison of speaking up behavior during conflict with real and virtual humans. *Computers in Human Behavior*, 52:12–21.

[Roldán-Álvarez et al., 2016] Roldán-Álvarez, D., Martín, E., García-Herranz, M., and 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.

[Rothman, 1990] Rothman, K. J. (1990). No adjustments are needed for multiple comparisons. *Epidemiology*, pages 43–46.

[Roussou et al., 2006] Roussou, M., Oliver, M., and Slater, M. (2006). The virtual playground: an educational virtual reality environment for evaluating interactivity and conceptual learning. *Virtual reality*, 10(3-4):227–240.

[Sabers et al., 1991] Sabers, D., Cushing, K., and Berliner, D. (1991). Differences among teachers in a task characterized by simultaneity, multidimensionality, and immediacy. *American Educational Research Journal*, 28(1):63–88.

[Santos et al., 2009] Santos, B. S., Dias, P., Pimentel, A., Baggerman, J.-W., Ferreira, C., Silva, S., and Madeira, J. (2009). Head-mounted display versus desktop for 3d navigation in virtual reality: a user study. *Multimedia Tools and Applications*, 41(1):161.

[Scherer, 2009] Scherer, K. R. (2009). Emotions are emergent processes: they require a dynamic computational architecture. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1535):3459–3474.

[Scherer et al., 2019] Scherer, K. R., Ellgring, H., Dieckmann, A., Unfried, M., and Mortillaro, M. (2019). Dynamic facial expression of emotion and observer inference. *Frontiers in psychology*, 10:508.

[Schubert et al., 2001] Schubert, T., Friedmann, F., and Regenbrecht, H. (2001). The experience of presence: Factor analytic insights. *Presence: Teleoperators & Virtual Environments*, 10(3):266–281.

[Schussler et al., 2017] Schussler, D., Frank, J., Lee, T.-K., and Mahfouz, J. (2017). Using virtual role-play to enhance teacher candidates' skills in responding to bullying. *Journal of Technology and Teacher Education*, 25(1):91–120.

[Schwind et al., 2017] Schwind, V., Knierim, P., Chuang, L., and Henze, N. (2017). " where's pinky?" the effects of a reduced number of fingers in virtual reality. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play*, pages 507–515.

[Sharples et al., 2008] Sharples, S., Cobb, S., Moody, A., and Wilson, J. R. (2008). Virtual reality induced symptoms and effects (vrise): Comparison of head mounted display (hmd), desktop and projection display systems. *Displays*, 29(2):58–69.

[Shernoff et al., 2020] Shernoff, E., Frazier, S., Lisetti, C., Buche, C., Delamarre, A., Gabbard, J., and Katherine, V. S. (2020). Evaluating the usability and instructional design quality of interactive virtual training for teachers (ivt-t). *Educational Technology Research and Development*, page in Press.

[Shernoff et al., 2018] Shernoff, E., Frazier, S., Lisetti, C., Buche, C., Lunn, S., Brown, C., Delmarre, A., Chou, T., Gabbard, J., and 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 et al., 2016] Shernoff, E. S., Frazier, S. L., Maríñez-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., and 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 et al., 2015] Shernoff, E. S., Lakind, D., Frazier, S. L., and Jakobsons, L. (2015). Coaching early career teachers in urban elementary schools: A mixed-method study. *School Mental Health*, 7(1):6–20.

[Shernoff et al., 2011] Shernoff, E. S., Maríñez-Lora, A. M., Frazier, S. L., Jakobsons, L. J., Atkins, M. S., and Bonner, D. (2011). Teachers Supporting Teachers in Urban Schools: What Iterative Research Designs Can Teach Us. *School psychology review*, 40(4):465–485.

[Slater and Steed, 2000] Slater, M. and Steed, A. (2000). A virtual presence counter. *Presence: Teleoperators & Virtual Environments*, 9(5):413–434.

[Slater and Wilbur, 1997] Slater, M. and Wilbur, S. (1997). A framework for immersive virtual environments (five): Speculations on the role of presence in virtual environments. *Presence: Teleoperators & Virtual Environments*, 6(6):603–616.

[Steuer, 1992] Steuer, J. (1992). Defining virtual reality: Dimensions determining telepresence. *Journal of communication*, 42(4):73–93.

[Straub et al., 2014] Straub, C., Dieker, L., Hynes, M., and Hughes, C. (2014). Using virtual rehearsal in tle teachlive^{TM} mixed reality classroom simulator to determine the effects on the performance of mathematics teachers. *2014 TeachLive National Research Project: Year 1 Findings*, pages 1–50.

[Su et al., 2019] Su, K.-W., Chen, S.-C., Lin, P.-H., and Hsieh, C.-I. (2019). Evaluating the user interface and experience of vr in the electronic commerce environment: a hybrid approach. *Virtual Reality*, pages 1–14.

[Sutcliffe and Gault, 2004] Sutcliffe, A. and Gault, B. (2004). Heuristic evaluation of virtual reality applications. *Interacting with computers*, 16(4):831–849.

[Sutcliffe and Kaur, 2000] Sutcliffe, A. G. and Kaur, K. D. (2000). Evaluating the usability of virtual reality user interfaces. *Behaviour & Information Technology*, 19(6):415–426.

[Sutcliffe et al., 2019] Sutcliffe, A. G., Poullis, C., Gregoriades, A., Katsouri, I., Tzanavari, A., and Herakleous, K. (2019). Reflecting on the design process for virtual reality applications. *International Journal of Human–Computer Interaction*, 35(2):168–179.

[Tcha-Tokey et al., 2016] Tcha-Tokey, K., Christmann, O., Loup-Escande, E., and Richir, S. (2016). Proposition and validation of a questionnaire to measure the user experience in immersive virtual environments.

[Tichon, 2007] Tichon, J. G. (2007). Using presence to improve a virtual training environment. *CyberPsychology & Behavior*, 10(6):781–788.

[Togelius et al., 2013] Togelius, J., Yannakakis, G. N., Karakovskiy, S., and Shaker, N. (2013). Assessing believability. In *Believable bots*, pages 215–230. Springer.

[Tracey et al., 2014] Tracey, M. W., Hutchinson, A., and Grzebyk, T. Q. (2014). Instructional designers as reflective practitioners: Developing professional identity through reflection. *Educational Technology Research and Development*, 62(3):315–334.

[Venkatesh et al., 2003] Venkatesh, V., Morris, M. G., Davis, G. B., and Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. *MIS quarterly*, pages 425–478.

[Vilhjálmsson et al., 2007] Vilhjálmsson, H., Cantelmo, N., Cassell, J., Chafai, N. E., Kipp, M., Kopp, S., Mancini, M., Marsella, S., Marshall, A. N., Pelachaud, C., et al. (2007). The behavior markup language: Recent developments and challenges. In *International Workshop on Intelligent Virtual Agents*, pages 99–111. Springer.

[Wages et al., 2004] Wages, R., Grünvogel, S. M., and Grützmacher, B. (2004). How realistic is realism? considerations on the aesthetics of computer games. In Rauterberg, M., editor, *Entertainment Computing – ICEC 2004*, pages 216–225, Berlin, Heidelberg. Springer Berlin Heidelberg.

[Webster-Stratton, 2005] Webster-Stratton, C. (2005). Teacher classroom management strategies questionnaire.

[Wenger, 1987] Wenger, E. (1987). Artificial intelligence and tutoring system: Computational and cognitive approaches to the communication of knowledge. *Califórnia: Morgan Kaufmann Publishers. Texto publicado na: Pátio-revista pedagógica Editora Artes Médicas Sul Ano*, 1:19–21.

[Whyte et al., 2015] Whyte, E. M., Smyth, J. M., and Scherf, K. S. (2015). Designing serious game interventions for individuals with autism. *Journal of autism and developmental disorders*, 45(12):3820–3831.

[Witmer and Singer, 1998] Witmer, B. G. and Singer, M. J. (1998). Measuring presence in virtual environments: A presence questionnaire. *Presence*, 7(3):225–240.

[Wu et al., 2014] Wu, Y., Babu, S. V., Armstrong, R., Bertrand, J. W., Luo, J., Roy, T., Daily, S. B., Dukes, L. C., Hodges, L. F., and Fasolino, T. (2014). Effects of virtual human animation on emotion contagion in simulated inter-personal experiences. *IEEE transactions on visualization and computer graphics*, 20(4):626–635.

[Zaidi et al., 2019] Zaidi, S. F. M., Moore, C., and Khanna, H. (2019). Towards integration of user-centered designed tutorials for better virtual reality immersion. In *Proceedings of the 2nd International Conference on Image and Graphics Processing*, pages 140–144.

[Zanbaka et al., 2005] Zanbaka, C. A., Lok, B. C., Babu, S. V., Ulinski, A. C., and Hodges, L. F. (2005). Comparison of path visualizations and cognitive measures relative to travel technique in a virtual environment. *IEEE Transactions on Visualization and Computer Graphics*, 11(6):694–705.

[Zanbaka et al., 2007] Zanbaka, C. A., Ulinski, A. C., Goolkasian, P., and Hodges, L. F. (2007). Social responses to virtual humans: implications for future interface design. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 1561–1570. ACM.

[Zibit and Gibson, 2005] Zibit, M. and Gibson, D. (2005). simSchool: The game of teaching. *Innovate: Journal of Online Eduation*, 1(6).

[Zielasko et al., 2016] Zielasko, D., Horn, S., Freitag, S., Weyers, B., and Kuhlen, T. W. (2016). Evaluation of hands-free hmd-based navigation techniques for immersive data analysis. In *3D User Interfaces (3DUI), 2016 IEEE Symposium on*, pages 113–119. IEEE.

VITA

ALBAN DELAMARRE

February 13, 1992	Born in Rennes, FRANCE
2016	M.Sc., Human Centered Computer Science National Engineering School of Brest Brest, France
2016	Engineering Degree National Engineering School of Brest Brest, France

PUBLICATIONS

1. Delamarre, A., Buche, C., and Lisetti, C. Modeling Emotions for Training in Immersive Simulations (METIS): a Cross-Platform Virtual Classroom Study. In *2020 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*, pages 78-84, 2020.
2. Delamarre, A., Buche, C., and Lisetti, C. A cross-platform classroom training simulator: Interaction design and evaluation. In *Proceedings of the 2020 International Conference on Cyberworlds*, pages 86-93, 2020.
3. E. Shernoff, K. Von Schalscha, J. Gabbard, Delamarre, A., S. Frazier, C. Buche, and C. Lisetti. Evaluating the usability and instructional design quality of interactive virtual training for teachers (IVT-T). *Educational Technology Research and Development Journal*, pages 1-28 ,2020.
4. Delamarre, A., C. Buche, and C. Lisetti. Aimer: Appraisal interpersonal model of emotion regulation, affective virtual students to support teachers training. In *Proceedings of the 19th ACM International Conference on Intelligent Virtual Agents*, pages 182-184, 2019.
5. Delamarre, A., S. Lunn, C. Buche, E. Shernoff, S Frazier, and C. Lisetti. Interdisciplinary collaboration and establishment of requirements for a 3d interactive virtual training for teachers. In *Proceedings of the 19th ACM International Conference on Intelligent Virtual Agents*, pages 185-187, 2019.
6. J. Clayborn and Delamarre, A. Living room conservation: a virtual way to engage participants in insect conservation. *Rethinking Ecology Journal*, 4:31, 2019.
7. F. Lasson, Delamarre, A., P. Redou, and C. Buche. A clinical decision support system to help the interpretation of laboratory results and to elaborate a

clinical diagnosis in blood coagulation domain. In *Proceedings of International Work-Conference on Artificial Neural Networks*, pages 109-122. Springer, 2019.

8. Delamarre, A. and J. Clayborn. Bridging the intention-behavior gap: Butterfly world 1.0. *John Pennekamp Annual Schaus' Swallowtail Butterfly Meeting*, 4 2019. Key Largo, Florida.

9. Delamarre, A. and J. Clayborn. The new frontier in effective butterfly and ecosystem communication and conservation: Virtual reality gaming frontier earth: Butterfly world 1.0. *North American Butterfly Association Miami Blue Chapter* meeting, 2 2019. Miami, Florida.

10. Delamarre, A. and J. Clayborn. Virtual reality and serious gaming: The new frontier in environmental education and insect conservation (presented in spanish). *XI Taller Sobre Ciencia, Tecnologia e Innovacion* 2019, 10 2019. Matanzas, Cuba.

11. E. Shernoff, S. Frazier, C. Lisetti, C. Buche, S. Lunn, C. Brown, Delamarre, A., T. Chou, J. Gabbard, and E. Morgan. Early career teacher professional development: Bridging simulation technology with evidence-based behavior management. *Journal of Technology and Teacher Education*, 26(2):299-326, 2018.

12. Delamarre, A. and J. Clayborn. The new frontier in pine rockland conservation: Virtual reality gaming: Butterfly world. *Pine Rockland and Tropical Botany Conference* 2018, 11 2018. Coral Gables, Florida.

13. Delamarre, A., C. Buche, M. Polceanu, S. Lunn, G. Ruiz, S. Bolivar, E. Shernoff, and C. Lisetti. An Interactive Virtual Training (IVT) simulation for early career teachers to practice in 3D classrooms with student avatars. In *FLAIRS 2017 - Proceedings of the 30th International Florida Artificial Intelligence Research Society Conference*, 2017.