Developing Effective Serious Games to Improve Cross-Cultural Competence

---

A Dissertation

Presented to

the faculty of the School of Engineering and Applied Science

University of Virginia

---

in partial fulfillment
of the requirements for the degree

Doctor of Philosophy

by

Brian Arnold An

August 2018

APPROVAL SHEET

This Dissertation
is submitted in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy

Author Signature: ______________________

This Dissertation has been read and approved by the examining committee:

Advisor: Donald Brown

Committee Member: Abigail Flower

Committee Member: Inki Kim

Committee Member: Gerard Learmonth

Committee Member: Shigehiro Oishi

Committee Member: ______________________

Accepted for the School of Engineering and Applied Science:

Craig H. Benson, School of Engineering and Applied Science

August 2018

UNIVERSITY OF VIRGINIA

DOCTORAL DISSERTATION

Developing Effective Serious Games to Improve Cross-Cultural Competence

---

A Dissertation
Presented to
the Faculty of the School of Engineering and Applied Science
University of Virginia

In Partial Fulfillment
of the requirements for the Degree
Doctor of Philosophy (Systems and Information Engineering)

by

Brian A. An

August 2018

Abstract

The recent globalization of modern businesses and the increased number of coalition-based military operations has necessitated that businessmen and soldiers alike be able to effectively communicate and interact with people of various backgrounds and cultures. Despite this recognized need and significant monetary investment, proven training and assessment methods for cross cultural competence (3C) continue to be lacking. The evolution of digital game-based learning systems introduces a new medium in which scalable training systems can be developed. This dissertation introduces and investigates novel frameworks and methodologies by which 3C games can be developed and players' performance can be assessed.

Although the contributions of this dissertation are specifically focused on serious games as they relate to 3C, the applications of these methods can be applied to various digital game-based learning systems (DGBL). The contributions of this dissertation can be organized into two categories 1) Procedural/Engineering contributions and 2) Scientific/Technical Discovery contributions.

Regarding the Procedural/Engineering contributions, this dissertation presents a novel taxonomy of 3C game research to identify gaps in current efforts and motivate the research presented in this dissertation. Additionally, this dissertation introduces the Cultural Dimension-Based Simulation Design Process. This process incorporates existing models of cultural differentiation to develop culturally-relevant dialogue systems for socially interactive DGBLs. The Cultural Dimension-Based Simulation Design Process is demonstrated in the development of the Chinese Cultural Dimension Training System(CCDTS), a game set in a Chinese University. The CCDTS was used in an experiment to determine whether the Cultural Dimension-Based Simulation Design Process was effective in improving 3C.

For the Scientific/Technical Discovery contributions, the author of this dissertation evaluates a novel eye-tracking based measure of performance(ETMP) for socially interactive DGBLs. This method involves the examination of player fixation duration on user inputs to evaluate the mental schema of the player's game selections. Additionally, the author of this dissertation evaluates the efficacy of the use of Head-Mounted Device (HMD) Virtual Reality (VR) compared to a traditional screen-based DGBLs for training cultural competency.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

3C	**C**ross **C**ultural **C**ompetence
AIC	**A**ikiki **I**nformation **C**riterion
AOI	**A**rea **O**f **I**nterest
ARC	**A**daptive **R**eadiness for **C**ulture
CCDTS	**C**hinese **C**ultural **D**imension **T**raining **S**ystem
CAMO	**C**ultural **A**wareness in **M**ilitary **O**perations
CMTS	**C**hinese **M**ilitary **T**raining **S**ystem
CQ	**C**ultural **Q**uotient
DGBL	**D**igital **G**ame-**B**ased **L**earning
DoD	**D**epartment **o**f **D**efense
EEG	**E**lectro**E**ncephalo**G**raphy
ETMP	**E**ye-**T**racking **M**easure of **P**erformance
HMD	**H**ead-**M**ounted **D**isplay
HCI	**H**uman-**C**omputer **I**nteraction
IDV	**I**n**D**i**V**idualism vs Collectivism
ISS	**I**ntercultural **S**ensitivity **S**cale
ITS	**I**ntelligent **T**utoring **S**ystem
IVR	**I**ndulgence **V**s **R**estraint
LREC	**L**anguage **R**egional **E**xpertise **C**ulture
LTO	**L**ong **T**erm **O**rientation
MAS	**MAS**culinity vs Femininity
MUVE	**M**ulti-**U**ser **V**irtual **E**nvironment
PDI	**P**ower **DI**stance
SID	**S**ocial **I**mportance **D**ynamics
SJT	**S**ituational **J**udgment **T**est
SL	**S**econd **L**ife
SLA	**S**econd **L**anguagel **A**cquisition
SME	**S**ubject **M**atter **E**xpert
UAI	**U**ncertainty **A**vo**I**dance
USMA	**U**nited **S**tates **M**ilitary **A**cadmey
VR	**V**irtual **R**eality

To my boys, Toby and Oakley.

may both of you grow up to explore, appreciate, and respect cultures other than your own.

To my wife, Jen, for her unwavering love, support, and dedication.

Chapter 1

An Introduction

"For him to have understood me would have meant reorganizing his thinking .... giving up his intellectual ballast, and few people are willing to risk such a radical move."

Edward T. Hall

In the United States, nearly every aspect of life requires some interaction with a person of a different background or culture than one's own. This is further exemplified when examining current military coalitions, international business partnerships, and recent political/governmental interactions. Despite this recognition, many recent incidents have led to catastrophic examples of cross-cultural miscommunication and lack of Cross Cultural Competence (3C) leading to political standoffs, failed business deals, and, in the worst cases, loss of life. In the context of military operations, recent military reports have identified that in operations in Iraq and Afghanistan, military forces "were not taught what local populations were really like" nor how to effectively deal with them[126]. While these opportunities for cultural interaction have existed for centuries, the theories used to characterize and quantify effective 3C are rather nascent. Furthermore, the application of these theories within modern instructional materials has recently gained significant attention with the advent of digital game-based learning (DGBL) systems. Given the cost-effective and repeatability of software-based products, these systems have enabled academic and commercial entities alike to develop high-fidelity scalable training systems. However, few researchers have evaluated the efficacy of these games. The aim of this dissertation can be divided into two general categories. First, the author of this dissertation explores *Procedural/Engineering Methods* to develop effective 3C games and, subsequently, presents methodological approaches and empirical evidence to support these contributions. Secondly, the author of this dissertation presents *Scientific/ Technical Discoveries* that contribute to the evaluation and presentation of 3C games.

For "how to develop effective 3C games," one must first examine the theoretical approach to 3C training. Bhawuk and Brislin examine two forms of training that they categorize as *experiential training* and *lecture training*[18]. The traditional method of lecture training is generally characterized by passive learning, defined problem solving, rational decision-making, and written responses, which have all been proposed to be inefficient at developing 3C[61]. Experiential training, on the contrary, is defined as a method whereby participants interact with people from various cultures to develop understanding and behavior that is conducive to effective cross-cultural interactions. Most modern 3C training systems have adopted entirely experiential philosophies or hybrid experiential/lecture-based systems.

Within the context of experiential training systems, one must also consider whether the experience is actual or virtual. Though actual experiences and interactions may be more effective than virtual ones, most institutions have targeted the use of virtual or computer-based training systems for cost and scalability reasons especially in the early phases of 3C education. The Department of Defense has adopted an evolutionary model of 3C education whereby actual experiences continue to play an important role in the improvement of 3C but only after a specific threshold of knowledge and performance has been achieved[47].

In the framework of a virtual environment, a consideration must also be made about whether the interactions within the virtual environment are made with real people or simulated people. This also has obvious cost and scalability implications though, in terms of efficacy, various researchers have made arguments for and found evidence to support both. In the case of real-person interactions in virtual environments, which is also referred to us puppeteering, systems such as the University of Florida's STAR simulator and the University of Virginia's Classroom Simulator have explored mediating real human interactions in mixed-reality systems[36]. Despite varying reports about their effectiveness, these systems face issues with scalability because both sides of the interactions require human control. On the contrary, simulated people or computer-driven avatars have gained popularity. However, the fidelity and complexity of these avatars widely vary.

Often, the design and delivery decisions related to 3C content are made based on resources and the impact of this on effectiveness has only been minimally explored. As such, the author of this dissertation explores and proposes design recommendations for 3C games specifically in the area of experiential virtual avatar-based DGBL systems.

As for how to measure 3C in DGBL systems, this question has primarily been explored in the area of cultural and cross-cultural psychology. The DoD has heavily invested in developing tools and training programs to improve 3C but the ways in which 3C can be quantitatively evaluated has not seen the same level of focus. Specifically, 3C measures that can be used to evaluate learning within DGBLs are lacking.

Survey instruments have been the primary tools with which 3C has been measured[99][25].

Although these tools have been widely used across varying domains and applications, their use as developmental measures has been limited. Given that many of the current instruments are self-reports, they often suffer from personal biases. Other tools, such as Situational Judgment Tests (SJTs), have been used as well[90]. SJTs typically involve participants reviewing a written scenario and selecting responses in a multiple-choice format. The responses selected correspond with varying degrees of correctness. Although these SJTs serve as effective developmental measures, they suffer from scalability issues because scenarios can only be presented to participants a single time without introducing a memory bias.

Recent technological advances have opened up new opportunities to measure performance specifically regarding physiological measures. Measures such as heart rate, Electroencephalography (EEG), galvanic skin response, and eye gaze have all been used to characterize human performance across a wide spectrum of skills and domains. Despite these advances, few researchers have investigated these measures within the context of DGBLs[120][72][37].

With regard to 3C measures, this dissertation explores a novel measure of quantifying performance in 3C DGBLs. Specifically, this dissertation presents an eye-gaze-based method of measuring performance that attempts to evaluate both decision-making processes and the actual decisions made within a DGBL system.

1.1 Organization of Dissertation

This research document is organized into the following sections:

- Chapter 2: This chapter provides background information relevant to the entire dissertation. Specifically, it includes descriptions of previous research that is related to 3C games and game development methodologies. Furthermore, this chapter includes a taxonomy of the state of the current research to identify gaps and shortcomings. Background information specific to each chapter is provided within the content of each chapter.

- Chapter 3: In this chapter, the author proposes the use of and explains a novel methodology for developing game content for 3C games. This method, called the Cultural Simulation Design Process, leverages existing research related to Cultural Dimension Theory. Furthermore, this chapter includes survey findings that are related to a specific model of Cultural Dimension Theory to substantiate its use in later phases of this research. This chapter is based on a conference article that was published after the 2017 International Conference on Social Computing, Behavioral-Cultural Modeling & Prediction and Behavior Representation in Modeling and Simulation.

- Chapter 4: This chapter provides the results of an experiment of game built using the Cultural Simulation Design Process. It includes an explanation of the process by which the CCDTS was created and a report on the results of an experiment whereby participants used the CCDTS and their performance was evaluated. This chapter is based upon a journal article that has been submitted to the *IEEE Transactions on Learning Technologies*.

- Chapter 5: This chapter investigates the effectiveness of the use of immersive systems specifically Head-Mounted Device (HMD) Virtual Reality(VR) in improving 3C. The development of an HMD VR based 3C game is described. This game is then used in an experiment to determine if the HMD VR 3C game results in an improved level of performance as compared to a traditional desktop version of the same game. The results of this experiment are reported. This chapter is based upon a conference article that has been accepted by the

2018 International Conference on Computer-Human Interaction Research and Applications.

- Chapter 6: This chapter includes an exploration of non-traditional means of evaluating performance in 3C DGBL systems. Specifically, it presents a novel methodology of evaluating performance using gaze behavior. The proposed method uses fixation duration along with the scoring profile derived from the Cultural Simulation Design Process to generate a composite weighted performance score. This score was recorded and evaluated through an experiment that was conducted with the CCDTS and the findings are reported. This chapter is based on a conference article that has been accepted for publication at the 2018 Winter Simulation Conference.
- Chapter 7: The final chapter of this dissertation includes a summary of the work that constitutes this dissertation and the significant findings. It also includes a description of the limitations of the described research as well as recommendations for future lines of investigation.

1.2 Summary of Research Contributions

This dissertation provides two categories of contributions related to 3C Games.

Procedural/Engineering Contributions:

- A literature review and taxonomy to identify gaps in the research on 3C games
- The development of a novel method called the Cultural Dimension Design Process for generating Cultural Dimension Theory dialogue content for 3C DGBL systems

- The demonstration of a game developed using the Cultural Dimension Design Process and statistical data that proves its efficacy
- The introduction and demonstration of the Role-Playing Situational Judgment Test to evaluate the learning benefits of the use of 3C DGBL systems

Scientific/Technical Discovery Contributions:

- The development of an ETMP in DGBL systems
- The demonstration of the ETMP in an experiment that shows the correlation of this novel measure to various other measures of performance
- The demonstration that HMD VR and traditional desktop variants of a DGBL result in insignificant performance differences
- The discovery that gender is a significant factor when determining learning outcomes in immersive DGBL systems

Chapter 2

Literature Review & Background Material

"If I have seen further than others, it is by standing upon the shoulders of giants."

Isaac Newton

This chapter begins with an overview of culture theory, cross-cultural competence, and cross-cultural training to provide a brief foundation for the theory behind this literature review. The background sections within the subsequent chapters, however, provide specific details that are relevant to content of each chapter including the specific cultural dimension frameworks and 3C knowledge from the field of cultural psychology.

In this chapter, the author then defines a taxonomy of 3C DGBL systems which is then used to motivate the research captured in this dissertation. It also includes a comprehensive literature review of the evolution of serious 3C games and their evolution into DGBL systems within the proposed taxonomy.

2.1 Culture

In an effort to understand what makes an individual cross-culturally competent, it is imperative to take a fundamental approach to understanding what differentiates cultures. Researchers in the field of Cultural Psychology have investigated and identified various frameworks by which cultures can be universally defined and differentiated. The seminal work of Kluckhohn's Value Orientation model first introduced the concept of universal core cultural variables or orientations and has led to the wide growth of research in the area[88].

Although many Cultural Dimension models have been proposed to date, the most widely accepted research in this area is the six-dimensional model of Geert Hofstede[94][68]. Through a factor analysis of over an initial survey population of over 100,000, Hofstede proposed a formula to normalize each of the six dimensions on a scale from 1 to 100[67]. These formulas served to calculate numeric scores for each dimension in the countries surveyed. After 30 years of critique and evaluation, this model and expansions of this model are the most utilized cultural dimension frameworks [113][42][91]. Hofstede characterizes culture across six dimensions: Individualism vs Collectivism(IDV); Power Distance(PDI); Uncertainty Avoidance(UAI); Masculinity vs. Femininity(MAS); Long-Term vs. Short-Term Orientation(LTO), and Indulgence vs. Restraint(IVR)[68]. Individualism is characterized by a society in which ties between individuals are weak and, thus, individuals will often look only look after themselves and their immediate families. On the contrary, a high score in collectivism reflects a society in which individuals are associated with protective and cohesive in-groups that expect obedience and the unquestioning loyalty of its members. Power distance is described as the degree to which the less powerful members of institutions and organizations within a country accept that power is distributed unequally [66]. Uncertainty Avoidance assesses the risk-tolerance and rigidity of a culture. Masculinity vs Femininity is generally described as the degree to which a society defines separate gender roles and distinct gender expectations. Long-Term vs. Short-Term orientation is characterized by whether a society adopts fundamentally Confucian values of one's outlook on life. Sixty-three countries have been evaluated across these dimensions.

Additionally, Trompenaar introduces a cultural dimension model that appears to be an expansion of Hofstede's model. Trompenaar's 7-dimensional model appears to overlap with Hofstede's model in many ways, but it also introduces several new dimensions including the Diffuse/Specific dimension and the Perception of Time dimension[134]. The Diffuse/Specific dimension can be described as the range of perceptions between a person's public and private space. The Perception of Time dimension captures whether a culture views time sequentially, as in discrete fixed units of time, or synchronously, meaning that the notion of time is abstract and flexible. Despite the face validity of Trompenaar's construct, our literature review showed that it did not share the same breadth of validation of Hofstede's model and, as such, was not considered in our proposed construct.

More recent research has questioned the validity of Hofstede's model (and others like it that have taken a relatively more fixed view of the factors that define cultures) under the premise that modern cultures can no longer be differentiated by ethnic and geographic lines as was the case in the past[77][122]. Despite these criticisms, a review of the relevant literature did not uncover a comparable model in terms of peer review and face validity.

2.2 Cross-Cultural Competence

As an extension to general Cultural Psychology or Cultural Dimension Theory, the field of 3C has emerged in an effort to characterize the qualities and measures of 3C. Thirty years of focused research has produced numerous theories and definitions of Cultural Competence. Despite a multitude of ideas, there is little consensus as to a standard definition or accepted measurement practice[26].

One of the earliest theories of "Intercultural Effectiveness" was developed by Hammer et al.[60] and is defined as a three-factor model comprised of (a) the ability to manage psychological stress, (b) effective communication, and (c) interpersonal relationships [26]. Research efforts to replicate the model were found to be highly variable [26].

More recently, Bennett's Developmental Model of Intercultural Sensitivity (DMIS) introduces a conceptual model of intercultural progression. Specifically, the DMIS defines an ethnocentric to ethnorelative scale by which one assesses his or her own culture with respect to other cultures [16][17]. Based on this theoretical model, the Intercultural Development Inventory (IDI) was developed as a means of assessing an individual's ethnocentrism/ethnorelativism based on the DMIS scale[99].

Later research conducted by Chen and Starosta defines Intercultural Sensitivity as an incorporation of sensitivity, awareness, and skills[24]. Sensitivity refers to the ability to "comprehend and appreciate cultural differences." Awareness refers to one's "ability to understand how culture affects thinking, behavior, and interactions." Skills refers to one's ability to communicate effectively in intercultural interactions. Specifically for the Sensitivity component, Chen and Starosta present the Intercultural Sensitivity Scale (ISS) which recent studies have shown ecological validity through modest correlations to 3C performance factors such as intercultural decision quality[99].

Recent research has focused on the role of metacognition as it pertains to Cultural Competence. Ang and Van Dyne propose the Cultural Intelligence (CQ) framework comprised of four components: Cognition, Metacognition, Behavior, and Motivation [8]. Ang and Van Dyne propose the CQ Inventory to measure each of these components with varying degrees of validation and replicability[26].

Specific to culture-general competence in military personnel, Rasmussen et al propose a 3C model referred to as the Adaptive Readiness for Culture(ARC)[115]. The ARC is categorized into the four broad domains of Diplomatic Mindset, Cultural Learning, Cultural Mindset, and Cultural Reasoning, and Intercultural Interaction. At the descriptive level, these four domains loosely correlate to the four factors of the CQ model though there are some notable differences. What's unique to this model as compared to CQ and previous models is its focus and genesis in deployed Department of Defense personnel. Additionally, the published validation of the ARC is based upon "semi-structured, incident-based interviews" which makes the ARC in its current form difficult to use as a developmental measure[115].

Despite the success of these 3C inventories and theoretical constructs in various applications, few have attempted and succeeded at using these inventories as measures of incremental development of 3C[99][21]. As the U.S. military and international corporations continue to invest in and deploy 3C training systems, more nuanced developmental methods of measuring 3C will be required to evaluate the effectiveness of these systems.

2.3 Cross Cultural Training

Cross cultural training has existed as early as the 1950s with Oberg's[110] research conceptualizing "culture shock" and the ways to remedy or prepare for the effects of it. Additionally, Hall's work proposing a taxonomy for cultural behaviors and identifying the cultural perceptions of physical space and time also served as seminal works in the area of cultural understanding and cross-cultural training[59][58].

Oberg characterized Culture Shock as a form of condition or disease given its physical and psychological symptoms, but these notions were later abandoned by researchers. What did remain was the understanding and acceptance that when individuals of one culture enter and live within the constructs of another culture, unfamiliar signs, symbols, and norms cause various levels of difficulty. Hall built upon this work and proposed a formal, informal, and technical approach to considering cultures[58]. Formal behaviors are considered those that are absolute, either correct or incorrect. Informal behaviors are those that are interpreted by observing the behaviors of other members of a culture, and in many cases, these behaviors can vary between members of a culture. Technical artifacts are considered those are transmitted through more explicit means such as written or codified rules. Unlike Oberg, Hall was able to operationalize his understanding of culture into a training program for individuals working abroad[59].

Since these foundational research efforts, many other researchers have investigated novel approaches to improving cross cultural competence. One of the most notable efforts is characterized by Triandis as the Cultural Assimilator[133]. In the Cultural Assimilator participants are presented with written scenarios involving two individuals of different cultures. The scenarios also called 'critical incidents' describe a conflict between the two individuals. Once a participant reads through the scenario, they are asked to reflect on the causes of the conflict or miscommunication and is then presented with various other behavioral options of varying degrees of appropriateness. Each behavioral option is accompanied by a description explaining the rationale for the option. This effort and others like it set the foundation for more experiential and sophisticated cultural training tools similar to those used today.

Technological advances have swiftly transitioned the content delivery of 3C training systems from paper-based to a variety of digital methods. The kinds of methods include everything from digital replications of written content to real-time agent-based conversations with semi-artificially intelligent avatars. Though the expansion and adoption of these various digital methods has been prolific, a literature review has shown that their evolution has been largely unstructured and has been difficult to examine progress with respects to the effectiveness of these systems. The following literature review of 3C DGBL systems attempts to structure these various research efforts into a logical taxonomy to characterize their current state as well as motivate future research.

2.4 3C DGBL Taxonomy

In order to properly contextualize the contents of this dissertation, it is necessary to describe the framework under which this work is derived and motivated. The proposed taxonomy identifies two factor sets of which the evaluation criteria are binned: **Development** and **Assessment**. The purpose of this specific taxonomy is to attempt to answer the question of "how well have various design and development factors known to be disparate among 3C DGBL systems led to more effective learning outcomes?"

The category of **Development** is defined as the set of game design and development factors that differentiate the 3C DGBL systems evaluated in this literature review. This is not intended to be comprehensive of all development factors that are considered in DGBL development, but strictly to those that have are disparate between the studies in this review.

The category of **Assessment** is defined as the set of factors that differentiate the type of game and/or participant assessment used in the specific studies evaluated in this review.

Development

Various developmental factors were identified as having been considered during the development of the 3C DGBL systems in this review. Given the lack of published research addressing efficacy of many of these factors, this taxonomy purposefully does not make any statement regarding the quality of these development factors but merely serves as a means to conceptualize the disparity between systems.

The development factors considered are as follows:

- Learning Method: The Learning Method factor is intended to identify whether a particular 3C DGBL was designed as an instructional or an experiential system. Early systems were predominantly instructional in nature often resembling lecture-style explanations of concepts and materials related to culture and 3C[61]. Recent studies have transitioned more towards experiential methods where learners are expected to participate and interact in actual conversations in order to develop both cognitive and behavioral components often associated with high 3C[25]. More sophisticated systems have found a medium between instructional and experiential systems whereby the experiences are augmented with Intelligent Tutoring Systems (ITS) or context aids which provide "out of body" explanations and rationale for various in-game situations and observations[90].

- Level of Immersion: The Level of Immersion factor addresses the level of realism and fidelity of the learning environment. Specifically, this targets the span of technologies available to deliver 3C content ranging from traditional desktop or fixed screen delivery systems to modern highly-immersive Head-Mounted Display (HMD) Virtual Reality (VR) systems. Given the novelty of modern VR, few studies have investigated the added learning value of modern VR delivery systems and this dearth has led to a burgeoning area of research in this space.
- Agent Fidelity: This factor addresses the flexibility and adaptability of the autonomous agents in the 3C DGBL. This can range from single string selection/response mechanisms all the way to complex agent-based models that incorporate behavior states based on cultural theory.

Assessment and Evaluation

Many of the studies reviewed reported some form of quality assessment of the 3C DGBL system in the study. The quality metrics used, however, span a wide range of methods and types which, in effect, makes it difficult to compare the results of various studies. In order to better examine the quality of proposed 3C DGBL improvements between studies, however, it is becoming necessary to make these comparisons and be able to draw conclusions that be used as the stepping stones for future research. In this particular taxonomy, the following assessment factors are described:

- Self-Report vs Independent: Many of the common 3C measures that exist in the literature are self-report surveys comprised of reflective assessments on one's beliefs and/or abilities[99]. Though these inventories have been used successfully in various applications, they have also been subject to scrutiny given their susceptibility to biases[25]. On the other end of the spectrum, studies have also attempted to use more independent methods of evaluating performance typically in the form of evaluative scores.

- Out-Game vs In-Game: Most of the 3C DGBLs evaluated in this review utilize an assessment method that is outside of the game environment. Generally these evaluations have been pre and post game assessments. It appears that the primary reason for this is that the assessments conducted in the studies in this review generally utilize existing inventories of 3C or are surveys to evaluate game experience. Other domains have begun to investigate in-game scoring methods in order to make an assessment while maintaining the sense of presence achieved by game environments.

- Explicit vs Implicit: Traditionally, explicit inputs interpreted as 'decisions made' have been used to evaluate performance in 3C DGBLs and most learning applications for that matter. The proliferation and cost-efficiencies achieved by physiological measurement devices, however, have opened up opportunities to make assessments implicitly based on these physiological responses. Until recently these devices were only available in academic and commercial laboratories, but as of 2014, consumer display devices are incorporating physiological measurements such as eye-tracking in order to augment user experiences. This additional data source has been used in various DGBL applications though no research was found with respect to 3C DGBLs specifically.

2.5 Review of 3C DGBL Systems

The above criteria and taxonomy are applied to seven unique research efforts that have generated and evaluated 3C DGBL systems to varying degrees over the past decade. Initially, each product and research effort is summarized. Following the summary is a categorization of the products research history based on the above taxonomy. Other research efforts listed at the end of this section were also discovered through the literature review though they lacked adequate detail in development or assessment to be considered in this taxonomy.

Elect BiLat

In 2006, through a collaboration between the U.S. Army Research, Development, and Engineering Command and the University of Southern California Institute of Creative Technologies, a cultural game called Elect BiLat was created[65].

The objective of this game was to allow soldiers to practice cultural engagements specifically in the case of bilateral negotiations. Though the story lines and characters are fictitious, they are derived from anecdotes of real situations that soldiers reported engaging in. The game objective is for participants to interact with in-game characters to achieve predefined objectives. As an example, a soldier may be required to meet with a local government official in order to convince him or her to change local laws, or to ask him or her about why a particular decision was made. Given that the game is designed specifically for customs and behaviors of Arabic countries, game participants are expected to navigate through customs, norms and various cultural expectations in order to meet their objectives.

The game is a first-person style game designed specifically for traditional desktop displays. The user interface is implemented as a dialogue selection menu screen (see Figure 2.1)[90]. In addition to the standard game dynamics, Elect BiLat is also built with an Intelligent Tutoring System (ITS) designed to provide in-game real-time feedback to user selections. This ITS is also developed with a model-scaffold-fade algorithm to reduce the available feedback as the user becomes more proficient[90].

Over the course of Elect BiLat's development, it underwent several experiments to evaluate various parameters of the game. In 2013, Lane et al evaluated whether fidelity results in a difference in efficacy between a 2D version of the game and 3D version of the game. They found that both versions performed equally. In the same study, they also evaluated whether conceptual or concrete feedback was more effective and found that conceptual feedback resulted in better learning transfer[90].

The primary method of evaluating performance is a pre/post Situational Judgment Test (SJT). This form of examination asks participants to review a scenario relevant to the teaching objectives in the game and then select the most appropriate response from a set of responses. Additionally, this study used an in-game score where the final interaction with the in-game character was without any ITS. Participants that made fewer errors in selecting behaviors were said to have performed better.

Second China Project

A team at the University of Florida developed the Second China Project as a means for participants to learn about Chinese culture and how to interact with Chinese nationals. The game is built using the Second Life gaming stack[140].

Variations of this game were created for the specific experiments that were conducted, but the primary game was developed as a Multi-User Virtual Environment

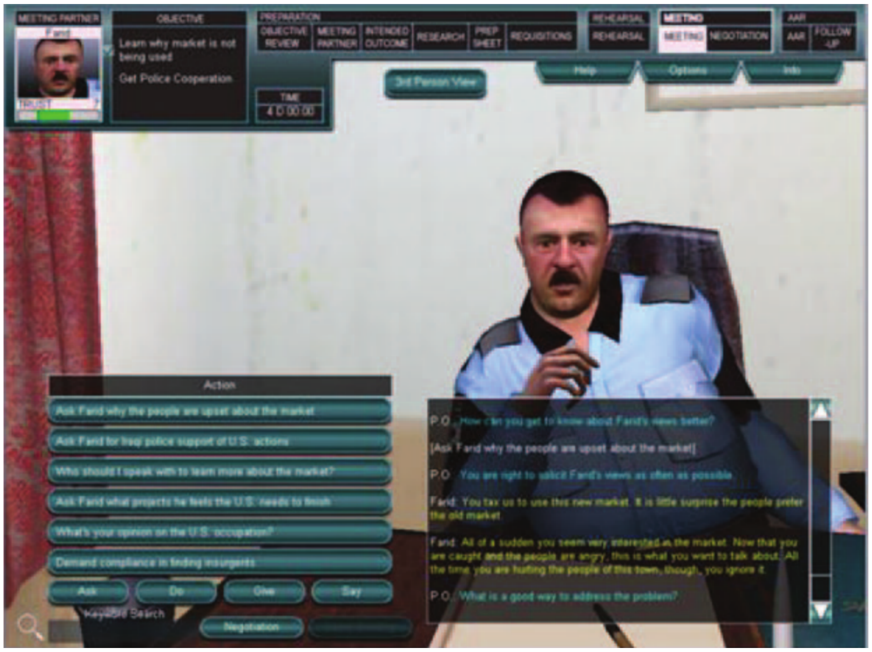


FIGURE 2.1: An in-game screenshot of Elect Bilat[90]

(MUVE) where participants navigate through multiple scenes. The game is played in a first-person perspective on a traditional desktop screen. The physical environments of the scenes include a Chinese city square, a tea house, an outdoor park, and a business office environment(See Figure 2.2). Artificially intelligent avatars also referred to as "bots" help the participant navigate through the game by providing instructional guidance as well as factual information regarding the specific scene in which the avatar resides. The avatars are driven by finite state machines though the nature and complexity of the states are not described in the published literature[28]. Given that the game is purely instructional, there is no means of evaluating performance within the game itself which lends us to believe that the game is entirely informative with little to no decision making by the player to exercise cultural competence. In addition to the primary version of the game, a web-based narrative version of the game was created for experimentation purposes[28].

FIGURE 2.2: An in-game screenshot of the Second China Project[28]

The effectiveness of the game was evaluated through a pre/post-test using a modified version of the Intercultural Sensitivity Scale (ISS) self-report Inventory developed by Chen and Starosta[24]. The specific modification changed the ISS from being being universal to all cultures to making it specific to Chinese culture. Participants were initially administered the ISS prior to being exposed to the Second China game as well as immediately following the completion of the game. Additionally, participants were again administered the same post test two weeks after the completion of the experiment to evaluate any decay or memory effects.

In their experiment, the primary Second China project game was tested alongside a less immersive web-based version of the game and found that the primary 3D version had a greater effect on 3C as defined by the ISS. In this study, they also found an interaction effect whereby those participants who were exposed to the 3D version of the game and were male scored significantly higher[28][29].

Cultural Awareness in Military Operations

A research group in Norway developed the Cultural Awareness in Military Operations (CAMO). The stated purpose of this research effort is to create a cost-effective and adaptive cultural game environment to train military personnel for operations abroad[114][132]. Unique to this effort is its emphasis on the execution of military operations in cultural diverse settings targeting a set of learning goals identified by the authors. The five overarching learning goals were identified as general military tactics, communicating with gender types, understanding religious influences, general social behaviors, and basic language skills[132]. The version of CAMO published in the literature focuses operations in Afghanistan comprising of what is referred to as 'mini-scenarios' each associated with 'cues' or "perceptual elements of the environment that influence the challenging decisions.[132]" Each mini-scenario is characterized by an interaction with a local person. The interaction involves an objective tied to each of the major learning goals of the game.

The game is developed within the Second Life gaming platform as a first-person perspective game intended for traditional desktop screen playing(See Figure 2.3). In order to achieve the stated objectives, each mini-scenario is designed as an interaction between the game participant and an avatar. In this case, the avatar is driven by a human role-player in order to facilitate dynamic and flexible conversations. The human role-players are provided with a scripted aid to provide some level of consistency between participants.

In terms of assessment, this research group conducted an experiment by having cadets of the Norwegian Military Academy participate in the game. The cadets played the role of the military forces entering the Afghan village while cultural experts played the role of the Afghan people within the village. The evaluation of the

FIGURE 2.3: An in-game screenshot of the Cultural Awareness in Military Operations project[132]

game was conducted two fold. First, the game was evaluated qualitatively through post-game discussions around the various conversations that occurred in-game. In this vein, the research group found that participants generally found the game to be relevant and informative[114]. Secondly, the participants were administered a Likert-scale based post-game survey to evaluate various aspects of the virtual game environment. Though not entirely conclusive, these results suggested increased understanding of Afghan culture and culturally effective communication. Additionally, participants found that more physical and perceptual realism would have increased the learning value of the game though there were no future efforts published in the literature exploring these findings.

Train for Virtually Every Locality

In collaboration with cultural expert Geert Hofstede[68], a research group developed a multi-tiered agent based simulation called Train for Virtually Every Locality (Traveller)[87]. The objective of Traveller is to provide an experiential pedagogical tool by which participants can learn about the affective and cognitive aspects of effective cross-cultural communication and interaction. Specifically, the affective goal intends to give users the ability to recognize emotional responses as possibly being cultural artifacts or derivations of cultural artifacts. The cognitive goal intends to provide cross-cultural context to similar actions. Since the user actions in a game are similar between games, participants are able to see how various actions drive different behaviors in avatars depending on the cultural context in which the action is taken[87].

The development and implementation of this game has evolved and branched in multiple research directions adding to the complexity and adaptability of the game. At its core, the game is designed as a first-person perspective game developed in the Unity3D game engine(see Figure 2.4). In the game, participants travel from country to country interacting with various individuals in what are referred to as 'critical incidents.' These 'critical incidents' are characterized by a specific objective that the player is trying to meet and must do so by selecting a sequence of physical behaviors and conversational responses. As an example, in one critical incident, participants must ask for directions from a group of strangers at a bar[87].

The avatars in the game are driven by a modular affective state machine called Fearnot Affective Mind Architecture (FAtiMA) developed by Dias et al[35]. At a high level, this state machine takes in an event(i.e. a specific interaction with a participant) and determines the impact of that interaction through an appraisal process. This appraisal process then determines the type and magnitude of the impact and updates the affective state of the avatar. The current affective state of the avatar is then used to select the behavior and response back to the human player. In order to incorporate a cultural component, a following research effort proposed the Social Importance Dynamics (SID) model to provide input into the FAtiMA state machine.

This model is based on Kemper's theory of Social Importance[82] which is summarized as the status that an individual places on another is related to the the degree to which someone is willing to perform "the wishes, needs, interests of the other[96]." in the SID model, this status is determined by cultural factors as prescribed by Hofstede's Cultural Dimension model[66]. As an example, cultures with high power distance would confer more Social Importance to those individuals who are more superior in the power hierarchy of that culture[96]. The quantitative dimensional scores developed through Hofstede's research serve as inputs to drive the status models of each agent[98]. Therefore, the simulation architecture of Traveller can be adapted to accommodate any national culture that has been quantified through Hofstede's dimensional research. Furthermore, synthetic cultures can also be developed and implemented through this by artificially assigning scores to each of the six Hofstede dimensions[98]. Other excursions of this research have investigated various verbal styles of the dialogue elements[97], gesture-based game interaction[87], and the incorporation of an empathic element to the FAtiMA model[121]).

The primary experiment conducted with Traveller focused on whether the agent models were positively viewed by participants from countries with comparable IDV dimension scores[98]. Participants were recruited from Portugal and the Netherlands specifically for the large difference in the IDV dimensions between the two countries as reported by Hofstede's research[68]. Specifically, Portugal was reported as the most collectivistic country in Europe whereas the Netherlands was reported as the most individualistic country in Europe. Participants were the randomly assigned to play either highly collectivistic or individualistic variations of the game. Following the game, participants were administered a survey about their interactions with the avatars in Traveller. The results from this experiment did not support their hypotheses and were largely inconclusive[98]. Despite a robust theoretical approach to

FIGURE 2.4: An in-game screenshot of the Traveller project[87]

developing culturally-specific avatars, no other research was found providing empirical evidence validating the game nor improving the cross-cultural competence of the game participants.

America's Army Adaptive Thinking and Leadership system

Alongside the John F. Kennedy Special Warfare Center and Schoolhouse(JFWSWCS), Sandia National Laboratories developed America's Army Adaptive thinking and Leadership system. This serious game was intended to supplement classroom-based instructional cultural training by giving participants the opportunity to practice culturally diverse interactions in a controlled environment. As such this research effort was born with the objective of developing a serious game to "exercise intercultural competence and metacognitive agility.[118]"

This game was intended to mimic real world situations of special operations forces in deployed locations and as such was developed in a Middle East setting. The premise of the game was a to resemble a military operation requiring intercultural negotiations and high-stress decision making. Participants were required to interact with non-US military forces, host country civilians, and various other non-US governmental personnel to achieve a pre-defined objective[118]. Aside from consulting with cultural experts, no other independent cultural or 3C theory was reported as having guided the development of the game content.

In terms of the game dynamics, this was developed as a first-person perspective puppeteering system whereby most of the virtual avatars were human-controlled(See Figure 2.5). Some of the lesser avatar roles were script-driven[118]. Participants played the role of the US forces in the area while cultural subject matter experts (SME) played the roles of the other human-driven avatars in the game. SMEs were provided with general scripts to follow in order to guide gameplay, but given the puppeteering aspect of the implementation, each iteration of the game varied based on the performance and decision-making of the participants.

The method by which participants were evaluated was by independent real-time Evaluator feedback. This method appears to entail the use of human raters that are are observing the game in real-time and are evaluating specific time-stamped interactions on a Likert-scale. Additionally, these raters also have the opportunity to record notes and comments based on these observations. The specifics of this method, however, were reported as being proprietary and the details which are not reported[116].

User experience of the game itself was also evaluated through a post-game questionnaire whereby participants subjectively judged the experience to be 'realistic' and 'engaging'[118]. As of this writing, no other evaluation of the game's efficacy was found in the literature.

FIGURE 2.5: An in-game screenshot of the America's Army Adaptive Thinking and Leadership system[118]

ICURA

Unlike the previous research efforts, ICURA is a serious game designed to teaching specific elements of contemporary culture and cultural etiquette instead of targeting skills associated with cross-cultural interactions[109]. Though having cultural knowledge is defined by several psychologists as contributing to general 3C, it is typically viewed as a single component of many that make up the various theoretical models of 3C[136][24]. As such, this game primarily targets the education of Japanese cultural artifacts and the assessment conducted evaluates whether participants attained the specific knowledge of the artifacts.

In ICURA, participants play the role of a tourist who has recently come to Japan(See Figure 2.6). The game is set in a fictitious Japanese town and the objective of the participant is to navigate the town in order to find a specific person. Through this journey, participants are presented with various cultural cues and artifacts linked to various subgoals.

The game is developed in the Unity3D engine as a first person perspective game. Aside from the participant, all the avatars are controlled by scripted dialogues. Additionally, the game includes an intelligent tutoring system that provides contextual information about cultural norms, interactions, and artifacts[53].

A formal evaluation was conducted on both the efficacy of ICURA as a pedagogical tool as well as its playability and usability. The evaluation on efficacy is conducted with a custom 12 question test which was administered before and after the game was played. Performance was measured by the difference of correct responses between the pre/post test. In their analysis, this research group found a significant increase in performance from playing ICURA. Regarding the playability and usability, a post-game survey found that participants found the game to be engaging and enjoyable though some participants did comment that the game controls were not intuitive[52]. In reflecting on the design and performance of the game, this research group found that utilizing multiple teaching methods (i.e. interactive and instructional) led to varying degrees of effectiveness of the game[53]. Specifically, the content that required the participant to interact with it actively resulted in greater performance gains.

Tactical Language and Culture Training System

Over the past decade, the Alelo Incorporation has been funded by various entities to develop cultural and language education software. One of these systems is the Tactical Language and Culture Training System(TLCTS). The TLCTS has multiple objectives in training 3C as it attempts to teach participants cultural and basic lingual skills simultaneously. This game combines classroom-like instructional lessons as

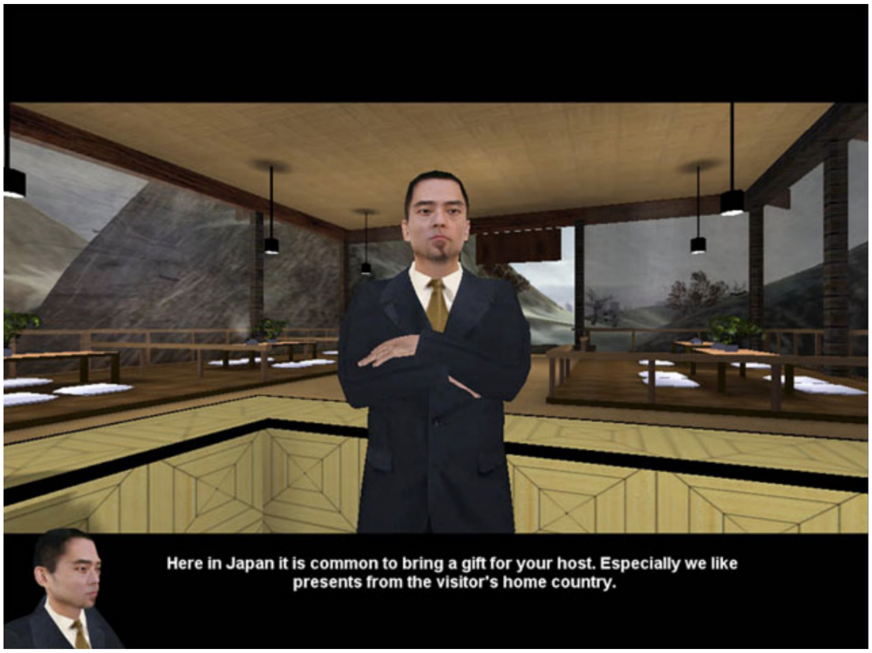


FIGURE 2.6: An in-game screenshot of the ICURA serious game[52]

well as an experiential serious game referred to as the Mission Game to practice the lessons taught in the instructional materials[74].

Several variations of this game have been produced to target specific languages and cultures to include Tactical Iraqi, Tactical Dari, Tactical Pashto, and Tactical French[74]. In the instructional component referred to as the Skill Builder, participants are presented with explanations and supporting materials for cultural artifacts. Within the Skill Builder component, participants are also able to practice basic linguistic skills that would later be used in the interactive game elements of TLCTS(See Figure 2.7). In the Mission Game, participants are expected to navigate a fictitious town and interact with locals for the purpose of achieving predefined objectives such as getting directions(See Figure 2.8). Unique to this research effort is the integration of natural language processing to ingest actual human utterances to interact with the avatars within the Mission Game.

The game is developed within the Unreal Game engine as a desktop screen game and has been programmed for a variety of inputs to include dropdown selections for non-verbal gestures, human utterance detection as well as common navigation through the game environment[75]. Despite the technical sophistication of the game dynamics, it appears that TLCTS is a platform by which subject matter experts and authors can build cultural content. As such, little is published regarding the content and the content development methods utilized to build the various TLCTS products[75].

Though publications relating to TLCTS reference in-game performance evaluations through quizzes and other mechanisms, the construct and mechanics of these methods were not found to be reported[75][74]. One study found that a majority of participants “had acquired a functional ability in Arabic” and gave the Tactical Iraqi TLCTS above average subjective ratings on a Likert-scale survey [74]. In an independent assessment of the game, Surface and Dierdorff found TCLTS to be effective in increasing Iraq Arabic speaking and listening skills though no reports an evaluation of 3C were made[129].

Other efforts

Other notable efforts were also discovered in this literature review though these studies lacked adequate detail in the available literature to be considered in this taxonomy. VECTOR, a cultural instructional game, places participants in a fictional town in order to achieve various in-game goals such as finding a bomber and attempting to stop him from hitting his next target[100]. Another product from the Alelo Corporation line of products, Virtual Cultural Awareness Trainer(VCAT) presents a curriculum which targets the knowledge, skills and attitudes (KSAs) of effective intercultural communication(See Figure 2.9). Similar to TLCTS, VCAT also includes an

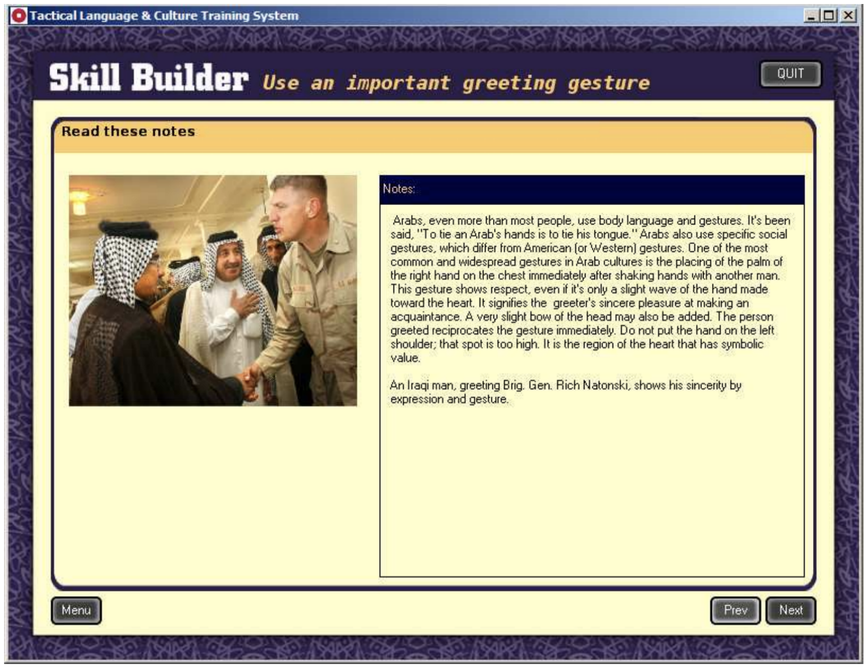


FIGURE 2.7: An in-game screenshot of the Tactical Iraq TLCTS Skill Builder[74]

FIGURE 2.8: An in-game screenshot of the Tactical Iraq TLCTS Mission Game[74]

instructional and an experiential component which work in tandem to meet specific

FIGURE 2.9: An in-game screenshot of VCAT[76]

learning objectives. Despite the relevance of VCAT to this literature review, there appears to be a lack of VCAT specific development details and any evidence of formal assessment/evaluations in the literature[76].

2.6 Summary of 3C DGBL Systems

Despite the multitude of research efforts in the area of 3C DGBL Systems, this literature review has shown a lack of organization and iterative direction with respect to these efforts. The purpose of this section is to organize and characterize several dimensions by which the previously described research can be categorized. In doing so, I hope to substantiate and motivate potential directions of research some of which are addressed in this dissertation. This taxonomy is by no means an attempt to draw conclusions regarding the efficacy of the various spectra defined below but

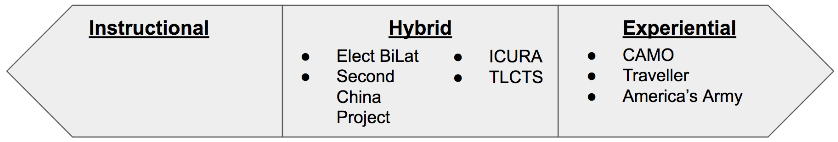


FIGURE 2.10: The Learning Methods used by the 3C DGBLs reviewed

an attempt to identify which factors have and have not been investigated. Additionally, the spectra presented are meant to serve as relative scales between the works presented.

Development

- Learning Method(Figure 2.10): Given the nature of the works investigated as well as the theoretical shift from instructional to experiential methods for 3C development[61], all of the works are either purely experiential or are hybrids between pedagogical methodologies. For the hybrid products, the main area of difference was between whether the instructional material was intertwined within the construct of the experiential component or whether the instructional and experiential components were entirely separate. The Second China project is an example where both methods are intertwined within the game environment and as such the participant is not removed from the immersive experience when the instructional content is presented[44]. On the contrary, the TLCTS separates the two environments such that the participant oscillates from an instructional mode where the cultural artifacts are presented and the Mission mode where the participant can experientially practice the target skills[75].
- Level of Immersion(Figure 2.11): Based on the time frame in which these games were developed, all of them were designed for desktop screen-based gameplay and, as such, this dissertation presents the first work where a 3C DGBL system

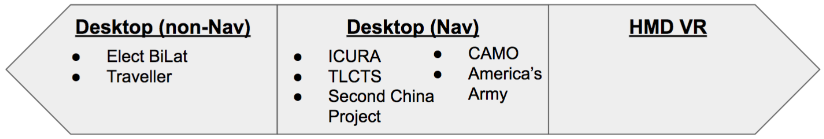


FIGURE 2.11: The Immersion Levels of the 3C DGBLs reviewed

is built for the HMD VR environment. Within the desktop environment, however, there were two levels of immersion that are identified. The first level can be described as non-Navigation meaning that participants are not in full physical control of their characters or avatars. In these systems such as Elect BiLat and Traveller, the game mechanics only allow participants to select conversational elements related to interactions. In the other games reviewed, participants are allowed to explore the environments in which they are located potentially allowing for a more immersive experience.

- Agent Fidelity(Figure 2.12): This factor is displayed on a single dimension though it may not necessarily be appropriate to consider along this spectrum. This factor attempts to capture the autonomy and realism of artificial agents. On one end of the spectrum, America's Army is developed with human-controlled avatars that allow for realistic and dynamic interactions though this methodology is not nearly as scalable as a scripted or model based architecture. Games such as Elect BiLat rely on selections from a finite set of inputs which result in a scripted set of known and predictable outputs[90]. Though this implementation is the most predictable and cost-effective to implement it is by no means dynamic and flexible as are real conversations. In an effort to expand the adaptability and fidelity of the interactions, the Traveller project uses agent-based state models to track an affective state of the agent as well as keep a memory of previous interactions in order to maintain a relationship between the avatar and the participant[98]. Though creating culturally realistic fully autonomous

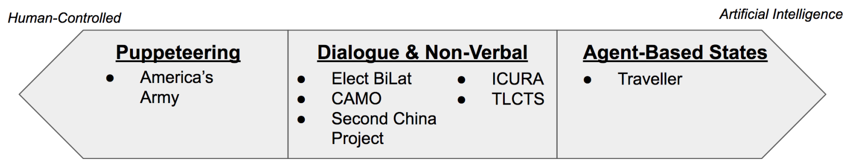


FIGURE 2.12: The Agent Fidelity of the 3C DGBLs reviewed

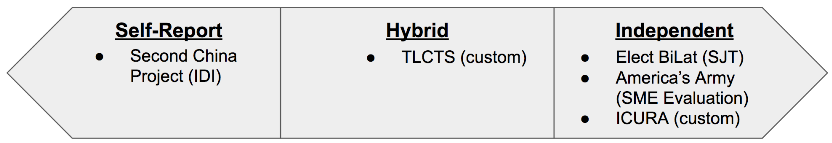


FIGURE 2.13: Whether the participants' performance were assessed by self-report or independent evaluation methods

conversational agents requires much more research, the work by Mascarenhas et al takes a step in that direction[98].

Assessment and Evaluation

- Self-Report vs Independent(Figure 2.13): In the games that were reviewed, it was found that various types of methods of evaluation were used to evaluate participant performance. In the case of Elect BiLat, pre/post Scenario Judgment Tests were used[90]. The Second China Project used a pre/post administration of the IDI[44]. America's Army used SME evaluations of individual performance[118]. For validation of the games, all of games reviewed relied on participant surveys to validate playability, usability and general sentiment towards the game.

- Out-Game vs In-Game(Figure 2.14: No games evaluated were found to have an entirely in-game assessment method to determine changes in 3C or related performance. The America's Army project did, however, use in-game subjective evaluations and annotations by SMEs of specific interactions which was

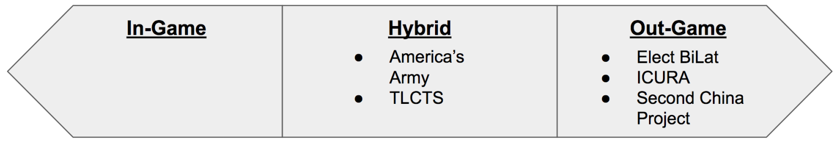


FIGURE 2.14: Whether the participants' performance were assessed by In-Game or Out-Game evaluation methods

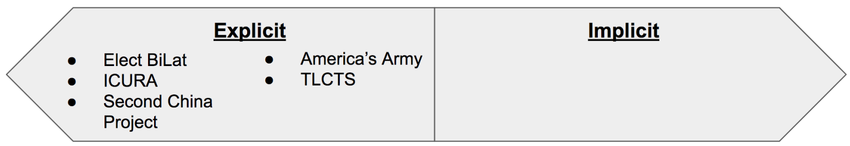


FIGURE 2.15: Whether the participants' performance were assessed by their explicit or implicit actions

then reported to the participants as part of qualitative feedback[118]. TLCTS also evaluated in-game responses as a part of their participant performance assessment though the specific methodology by which that was conducted was not reported in the literature[75].

- Explicit vs Implicit: All the games reviewed assessed participant performance through some form of explicit measure of performance. This is to say that what was assessed was the outcome of a decision that was made by the participant whether it was a decision by the participant to respond to an avatar or by a participants self-evaluation of performance. The America's Army project, however, did mention that the decision-making processes were discussed in the after-action review sessions, though they were not explicitly evaluated[118].

Areas of Research

Through this review and taxonomy of existing research and the taxonomy under which they were evaluated, several notable areas of needed investigation are apparent. Specifically in the area of game development, investigation into whether highly immersive display media have an effect on the quality of the learning experience in 3C DGBLs. Other domains have found positive results when teaching various skills and topic in more immersive environments[4][141]. In the area of DGBL assessment methods, in-game and implicit measures of performance are areas of investigation for future research. Hypothetically, in-game measures could provide insight into developmental performance such as learning rates. Additionally, with the proliferation of physiological measurement devices, it would seem worthwhile to investigate the use of implicit measures of performance to gain better understanding of 3C related cognitive processes and decision making.

In summary, this literature review highlights a number of areas that have yet to be explored in the area of 3C DGBL systems and motivates some of the research presented in this dissertation.

Chapter 3

A Design Process for Cultural Dimension-Based Simulations

This chapter is based on an article published as a conference article at the 2017 International Conference on Social Computing, Behavioral-Cultural Modeling & Prediction and Behavior Representation in Modeling and Simulation.

Cross-Cultural Competence (3C) has been identified by both corporate and military establishments as a necessary requirement for success in both modern business and military operations [55][108]. DoD Instruction 5160.70, Management of the Defense Language, Regional Expertise, and Culture (LREC) program, identifies regional expertise and culture as mission critical competencies for the Defense LREC program. Specifically, the Department of Defense (DoD) has recommended the creation of a developmental model for 3C expertise that prescribes the progression of competency development [101]. For this reason, we developed an objective and quantifiable means of measuring cultural competence through the adoption of Cultural Dimension Theory and an avatar-based simulation. This chapter describes exploratory survey results to identify the applicability of Cultural Dimension Theory. Additionally, this chapter discusses the Cultural Simulation Design Process which incorporates Cultural Dimension Theory in a simulation intended to measure 3C.

3.1 Cultural Dimension Theory Applicability

In an effort to determine the applicability of Cultural Dimension Theory as a means of measuring the Cross-Cultural Competence of DoD personnel, we conducted a perspective-taking survey of United States Military Academy (USMA) cadets. Seventy-four USMA cadets whom had recently returned from semester exchanges were surveyed. Given the extensive language and cultural training that the cadets had received in preparation for their exchange programs as well as during their time abroad, our intention was to determine if the cadets identified differences between their own cultural and that of their exchange program consistent with the findings of Hofstede's research[67].

Objective

The objective of this survey was to determine whether the cadets' perceived cultural differences were consistent with the published Hofstede's Cultural Dimension studies of the same countries. Similar methods have been used in other cross-cultural studies with mixed results though no studies were found to have used the Hofstede's Values Survey Module 2013 Inventory in this fashion[64].

Method

All the cadets were administered Hofstede's Values Survey Module 2013. Once complete, the cadet's were immediately administered the same survey with the instruction to answer the questions as they would expect the people in their exchange country to respond. The cadet semester exchange countries were concentrated in five countries (Taiwan, France, Germany, Mexico, Brazil). All the cadets designated

American as their primary culture. They were not given any descriptive information about the specific dimensions.

Results

Fig 3.1 depicts a comparison between the difference of means observed in this study and the difference of means observed in the published Hofstede results of the same countries. The scales of each of the dimensions from the higher values to the lower values are as follows: High to Low Power Distance, Individualistic to Collectivistic, Masculine to Feminine, High to Low Uncertainty Avoidance, Long-Term to Short-Term Orientation, Indulgent to Restraint. The numeric scale in Fig 3.1 is the absolute difference between the mean score of the USMA cadets and the mean scores as published in previous Hofstede publications. This is intended to show whether the observed difference of means in this study trend in the same direction as Hofstede's results.

Discussion

In examining Fig 3.1, twenty-four of the thirty calculated cultural dimensions spanning the five countries trended in the same direction as that of Hofstede's published results. However, a couple notable dissonances were observed as described in this section. Though several hypotheses are proposed below for these dissonances, we are unable to definitively explain the root cause without further investigation.

In the case of Taiwan, we observed that the survey participants' perceptions of Taiwan were generally more masculine than feminine in contrast to the more neutral perspective reported in Hofstede's work. Other studies have replicated the result

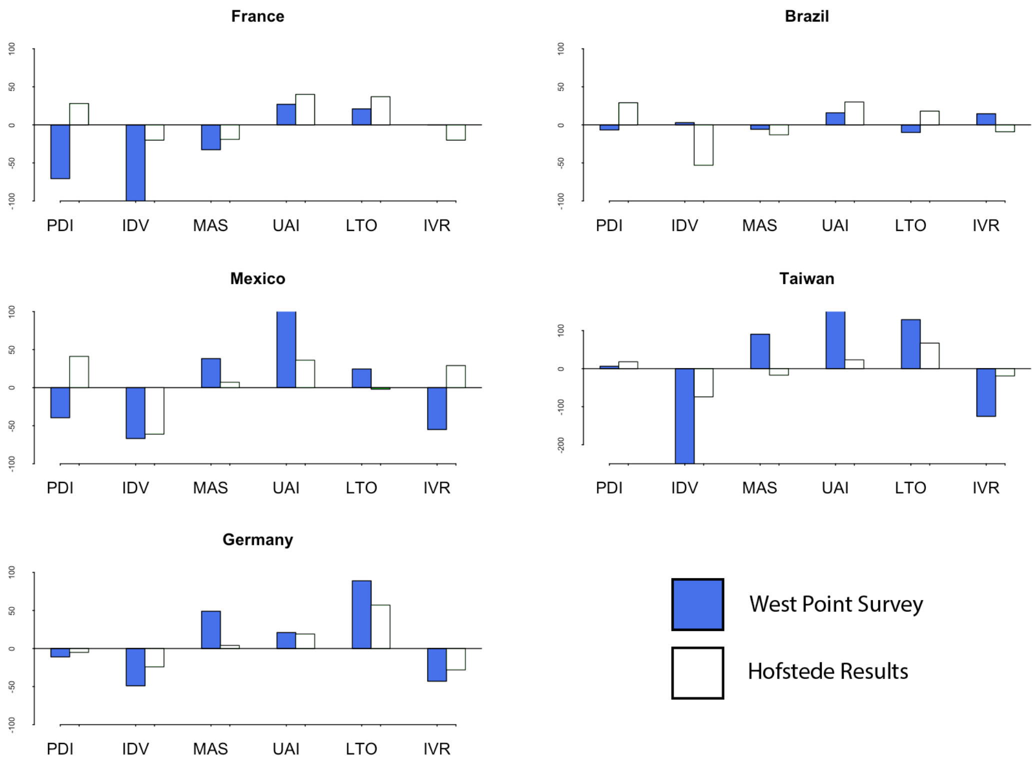


FIGURE 3.1: Comparison of USMA results and Hofstede results for mean difference

we observed [146][43]. Wu et al. explained the difference due to the rapid change in the U.S. workforce and gender roles since Hofstede's results were published which in turn would change the comparative result.

Another notable difference was in Power Distance for France. Through discussions with the participants, it was noted that they observed a less rigid rank structure among the French military than what they were accustomed to at USMA. This may be more representative of French military culture rather than French culture as a whole.

This result lends some evidence that the cultural dimensions can be correctly perceived which presumably would allow an individual to better interpret behaviors and respond accordingly. To provide further support to this conclusion, we plan to correlate these findings to established self-report measures of 3C.

3.2 Cultural Simulation Design Process

Previous iterations of this research[107] and other related efforts have not attempted to incorporate Cultural Dimension Theory into their simulation design processes. As such, these simulations failed to present the users anything beyond rudimentary or superficial cultural interactions. As an evolution of the previously developed avatar-based cultural training simulation[107] and the results of the USMA study, this research effort systematically incorporates the previously discussed Cultural Dimension Theory into the simulation design process to address DoD requirements[101].

Overview

The Cultural Simulation Design Method extends existing Human-Computer Interface (HCI) design principles as described in other simulation design efforts[117]. The specific addition uses Cultural Dimension Theory to influence the development of the branched dialogues in order to increase the efficacy of the pedagogical value and assessment of the simulation. In doing so, it presents players of the simulation to inductively learn the existence of the cultural dimensions which we hypothesize can increase the players Cross Cultural Competence.

Dimension Selection

In order to determine which Cultural Dimensions could be used to evaluate cross-cultural competence, we first determined the participant's most affiliated culture as well as the target assessment culture of the simulation. As an example, a participant considers himself most closely aligned with American culture and the simulation is designed to reflect Chinese culture.

We then determined which Cultural Dimensions were the most different between the participant's culture and the simulation culture. Hofstede's previous work shows that Power Distance, Individualism, Long-Term Orientation, and Indulgence have the most notable differences[66]. From these remaining dimensions, a subset was selected as the target dimensions in the simulation.

Dialogue Generation

In-game branched dialogues with avatars are the sole means of progressing through the simulation. These dialogues are contextually developed based on a predetermined storyline.

In response to an avatar, participants must select a response from a list of predetermined responses. The spectrum of these responses is crafted to capture the spectrum of the target cultural dimension. This methodology is repeated throughout the entirety of the dialogue trees in order to maximize the number of times a cultural dimension is exposed to the participant.

Cross-Cultural Competence Assessment

Given the integration of Cultural Dimensions into the dialogue structure, cross-cultural competence is assessed based on the number of culturally appropriate responses selected. We hypothesize that one's ability to recognize and respond to the appropriate cultural dimensions is reflective of a higher cross-culture competence.

Simulation Development

Using Unity3D as the simulation development engine and the previously described design process, we developed a five scenario simulation set in a Chinese university. The initial background model was purchased from the Unity3D marketplace and was subsequently tailored for our storyline. The avatars were each individually developed in Blender and Mixamo Fuse. Using the Pixel Crushers Dialogue System, we created branching text/voice dialogue trees.

3.3 Conclusion

This project builds upon previous work[107] to assess and improve a soldier's cross-cultural competence through the use of avatar-based simulation systems.

The Cultural Dimension survey of USMA cadets provides initial validity to the use of Cultural Dimensions to measure cross-cultural competence. Additionally, the novel Cultural Simulation Design Process introduces a systematic methodology to develop cultural simulations grounded in current theory.

Chapter 4

Development and Evaluation of a Cultural Game

This chapter is based on an article submitted to be published as a journal article in the IEEE Transaction for Learning Technologies Journal.

Modern globalization has made possible an unprecedented potential for international communication and cooperation between policymakers, militaries, and businesses. As a result, it has become increasingly important for effective politicians, soldiers, and businessmen/women to acquire Cross-Cultural Competence (3C) defined by the Department of Defense (DoD) as the “set of knowledge, skills, and affect/motivation that enable individuals to adapt effectively in cross-cultural environments[55].” As such, the DoD and many multinational corporations have spent considerable resources developing methods to train and assess the 3C of their personnel with varying degrees of success.

These investments have led to various training methods and developmental measures. The field of Cross Cultural Competence Psychology has surfaced various inventories though research in developmental measures is in its infancy.

The training methods have continuously evolved though one method that has gained significant traction due to its cost effectiveness and recent technological sophistication has been virtual immersive serious games. Virtual immersive serious games have long been used for training and assessment purposes for military personnel in combat situations as evidenced by several commercially available simulation development platforms such as Virtual Battlespace [107]. The agility and adaptability of these systems have successfully been used to replicate physically demanding and emotionally stressing situations in order to prepare soldiers for combat situations. While many research efforts have identified frameworks to address various aspects of game development, few have proposed dialogue creation frameworks specific to the area of cross cultural competence[116].

This study presents a novel framework to generate culturally relevant content and dialogue for the purpose of teaching cultural dimensions and culturally appropriate communication. This framework leverages fundamental cultural differences explained by Cultural Dimension Theory defined by Geert Hofstede as the "effects of a society's culture on the values of its member, and how these values relate to behavior[68]." By exposing participants to the spectrum of a cultural dimension instead of explicitly teaching appropriate behaviors, we hypothesize that participants will be better equipped to generalize their training to situations beyond those specifically presented to them. Though some studies have incorporated various degrees of Cultural Dimensions into game designs, we have not found any studies that explicitly tie Cultural Dimensions to each dialogue interaction as shown in this study[98]. We present the results of an evaluative experiment of a game developed using this framework showing that this framework has the potential to be used to inform the development of cross-cultural content for the purpose of higher-level 3C education and assessment.

4.1 Background

Virtual Immersive Game Development

Since the introduction of computer-based games, researchers have attempted to codify game design methods and best practices to achieve the greatest user experience. Early researchers developed high-level theoretical frameworks to achieve a "sense of immersion" or "sense of presence" in virtual environments [123].

As gaming media and the technological sophistication of gaming environments has evolved, it has become necessary to refine the design process specific to the domain in which it was being applied. With the formalization of serious computer-based games, a wide range of studies have been conducted to evaluate the efficacy of these games and, subsequently, develop specific guidelines and frameworks to improve the gaming and pedagogical experience[104][9][95][106]. Each of these frameworks proposes different approaches to game design, but fundamentally all the guidelines and frameworks revolve around increased immersiveness, adaptive complexity, and relevant gameplay/interactivity. As useful as these frameworks are to serious game designers, researchers have continued to increase the specificity of design methods to game genres.

In the area of socio-cultural games, an inherent design consideration is the way in which culture is personified and exemplified in the in-game characters and the virtual environment. In the case of serious games targeting factors related to 3C education and growth, this becomes increasingly important. Several research groups have identified guidelines to develop culturally accurate persona and interactions. Raybourn introduced the "Designing from the Interaction Out" and a subsequent evolution to this method called the "Simulation Experience Design Framework"[116][117].

These methods emphasize four factors that should be considered together to engender a culturally rich experience. These four factors are (1) the type of communication (2) the physical place or context (3) the narrative/scenario and (4) the emergent culture as a result of the communication. Despite the particular relevance to the socio-cultural domain of serious games, this framework lacks specificity regarding the implementation of cultural factors in order to recreate culturally accurate and enriching gaming experiences.

Specific to the virtual agents, various researchers have focused on both cognitive models and behavior models to explain and define an agent's in-game responses and interactions.

Jan et al. propose a gaze and proxemics-based model addressing specific cultural differences [71]. Focusing on verbal communication, Mascaranehas et al. investigated the use of verbal style and tonality to exemplify cultural differences [97]. Specifically, the authors generated video clips of a dialogue sequence spoken in various verbal styles exemplifying various Hofstedete Dimensions. Participants were asked to rate each of the video clips for cultural appropriateness. The results showed mixed agreement between the culture of the participants and the cultural dimensions exemplified in the video clips.

Culturally specific cognitive models, however, have drawn much more research attention. For decision-making and negotiation, the Culturally-Affected Behavior (CAB) model ties explicitly coded cultural norms and behaviors to a task-oriented model[127]. The CAB model leverages a Theory of Mind implementation wherein the human participant ascribes a belief model to the avatar to whom he is interacting. As a user interacts with the avatar, each interaction updates the task probability model and provides feedback to the user through positive or negative responses.

More recently, Mascarenhas et al. proposed the Social Importance Dynamics (SID) model which operationalizes Kemper's Status Power Theory, the notion that all social activity is motivated by status and power[82][98]. In this model, all actions claim or confer social importance between agents. The magnitude of the claim or conferral is determined by a cultural modifier that is determined by Hofstede's Cultural Dimension Theory.

Gender and 3C

Though 3C has generally been considered a skill set attainable by both genders, studies have shown that females report higher 3C or more often exhibit personality traits consistent with higher 3C [34][69] [73]. Many of these studies show single snapshots in time though recent studies begun to report similar gender differences in longitudinal developmental studies with mixed results[29][19][7]. Given the body of work suggesting that females either report higher 3C or exhibit relevant personality traits, we also hypothesize a gender difference in 3C growth derived from game performance.

Inductive(Implicit) Learning

Learning is considered deductive or inductive when learners do or do not receive information concerning the foundational rules of inputs, respectively[39][143]. In the fields of education and cognitive psychology there continues the debate of whether inductive or deductive learning is more effective.

The body of work with respect to inductive and deductive learning of culture or 3C is extremely limited and, as such, we turn to the research associated with Second Language Acquisition(SLA) given its relationship to 3C[112]. Reber's seminal work in learning an artificial language investigated the effects of both explicit(deductive) and implicit(inductive) learning and found that implicit learning was more efficient for complex concepts. Additionally, Reber found that implicit learning leading to tacit knowledge resulted in more accurate decisions when novel stimuli were introduced and had much less individual variability as compared to explicit learning[119]. More recent works have also built upon Reber's work finding that implicit or inductive learning can improve performance and retention though research in this area is far from definitive [138][41]. This area of research continues to be extremely active as other research efforts have also found contradictory or mixed results[23][131].

Given the promising findings in inductive learning in SLA, however, we have chosen to adopt this learning methodology in our framework.

4.2 Cultural Dimension Design Process

We built a game design and development framework that leverages Hofstede's Cultural Dimensions in each user interaction in order to facilitate 3C development. Researchers have indicated that this may be foundational to cross-cultural competence though few have attempted to systematically integrate them into a serious game[48][21].

Overview

Our framework extends existing Human-Computer Interface(HCI) design principles as described in other simulation design efforts[117]. Importantly our framework uses Cultural Dimension Theory to influence the development of the branched dialogues in order to increase the efficacy of the pedagogy and assessment of the simulation. The framework is detailed below:

Select Target Training Culture and Trainee's Culture

We first determine the participant's most affiliated culture as well as the target training culture of the simulation. As an example, a participant considers himself most closely aligned with American culture and the simulation is designed to reflect Chinese culture. A problem with many cultural serious games today is that they portray the training cultures in the same way regardless of the trainee's background. Studies have shown, however, that cultures are viewed uniquely through the lens of the person who is viewing a foreign culture[68][66]. Therefore, it would seem prudent that 3C education systems target the values with the greatest disparity between the trainee and the target assessment culture.

Identify Discriminating Cultural Dimensions

Subsequently, we use existing Cultural Dimension research to identify the most discriminating Cultural Dimensions between the Target Training Culture and the Trainee's Culture. Several lines of Cultural Dimension research can serve as the foundation for this determination but for our research, we use Hofstede's framework given its extensive history and empirical validation [94][68]. Hofstede proposes a formulation that normalizes the dimension scores of each country on a scale from 1 to 100[67]. Using this scale, many research groups have used various thresholds to determine discriminating thresholds though several studies have used an approximate difference of 40 points to be considered a large difference[105][147]. For this reason we used a threshold of a 40 point difference to determine discriminating differences between countries. As an example, in the case of a simulation being designed for American trainees learning Chinese culture, Hofstede's research has shown that Power Distance, Individualism, Long-Term Orientation, and Indulgence are the most discriminating cultural dimensions[66]. Therefore, these dimensions were implemented into the simulation.

Derive Context and Game Situation

The Cultural Dimensions alone are not adequate to develop a game situation that captures these dimensions in a contextually relevant way as they represent abstractions of cultural differences. Therefore it is important to identify and contextualize real behaviors where these dimensions could manifest. This can be accomplished in a variety of ways though being in a university setting, we were able to conduct focus groups of subject matter experts (SME) consisting of foreign language professors, exchange students, and linguistic experts. Participants were asked about culturally difficult situations or stress points in their home countries or in abroad experiences. Additionally, they were asked to hypothesize any underlying reasons for these experiences and relate them to specific Cultural values or dimension. As an example, the derived context of high respect for positions of authority (e.g. professors and elders) was most often linked to Power Distance as the underlying Cultural dimension.

Develop an Adaptive Storyline and Dialogue Trees

In general, developing a believable interactive story is critical to achieve a sense of place and elicit realistic user behaviors[33]. This is no different for this framework, though this step provides specificity as to how this story needs to be developed to achieve the pedagogical value of a cultural serious game. The story line and plot must be culturally relevant. Ideally, this relevancy can be established through focus groups or direct access to SMEs in the previous phase, but, at a minimum, through a thorough review of the literature[118].

Through thoughtful consideration and design, the storyline allows for a set of interactions and experiences across the spectrum of the Cultural Dimensions that are being targeted. Specifically, the user dialogue options must be deliberately created to span the spectrum of a Cultural Dimension. The purpose of this is to give users the opportunity to inductively learn a dimension. This would then allow the user to extend this knowledge to other instances of dialogue where a similar dimension is presented. Figure 6.2 shows an example of a Chinese greeting. In this example, the greeting options are presented across the spectrum of the PDI dimension with the far left being the highest PDI and the far right being the lowest PDI.

Replicating this spectrum through the dialogue is critical to giving the user several opportunities to intuit the existence of this dimension and ultimately to use this knowledge in novel situations. The number of times that it is replicated is dependent on the number of dimensions being addressed in the serious game and the total duration of the story line.

Finally, it is highly recommended that these dialogue trees are made in the native language of the culture that is being targeted. This, obviously, is dependent on the linguistic skills of the trainee population, but in many societies, previous research has shown that culture and language are inseparable and thus conveying a value system is most effectively done in the native language[63]. See Appendix A and Appendix B for full scenario dialogue trees.

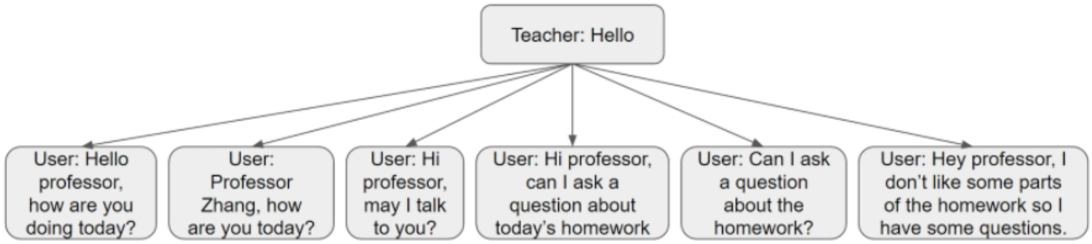


FIGURE 4.1: Greeting dialogue options across the PDI dimension for a 3C game for Chinese culture (English Translation)

Evaluate and Iterate

Given the subjective nature of the development of the dialogue trees, an independent evaluation step is imperative. This evaluation step is intended to validate the degree of to which each dialogue option is considered to be appropriate or inappropriate within the context of the targeted Hofstede dimension. This can be achieved in a variety of ways and will likely be dependent on available resources. We have chosen to use interrater reliability, κ, as a means of assessing dialogue validity[125] .

Chinese Cultural Dimension Training System (CCDTS) Development

Using Unity3D as the simulation development engine and the previously described framework, we developed a five scene simulation placed in a Chinese university. The initial background model was purchased from the Unity3D marketplace and was modified for this use case. The avatars were each modelled and developed in Blender and Adobe Mixamo Fuse. Figure 4.2 shows a depiction of the Classroom scene.

Using the Unity3D plugin Pixel Crushers Dialogue System, we created a branching text and voice dialogue tree for each scene. The addition of the recorded voice to the dialogue trees was critical given the importance of tone in the Chinese language. In each scene, the player was required to navigate a conversation with a non-player character through the selection of responses in the branching dialogue tree. In order to mitigate against the likelihood of correctly guessing responses, most of the dialogue sets included 4 to 6 options to select. To provide a sense of scale, the dialogue tree for each of the five scenes ranged from 15 to 25 unique interactions. The approximate play time for the entire game was estimated to be 45 minutes.

In order to support an equal weighting of each cultural dimension to the final score, our dialogue trees were created so that each dimension was presented to the player an equivalent number of times. In doing so, the final game score of the participant represented an equally weighted aggregation of each of the dimensional scores of the participant.

The storyline of this game is to complete a student project and work with peers and a professor in a Chinese university setting. The group project is assigned during class, discussed over lunch and in the library, and finally the student visits office hours to clarify questions individually with their professor. Key interactions include addressing new people, discussing American and Chinese stereotypes, respectfully interacting with authority figures, navigating conversations with peers on uncomfortable topics, and scheduling social gatherings with familiar and unfamiliar classmates.

We also incorporated a feedback system into the first three scenes to provide feedback to the participants on their dialogue choices. This feedback system provided cultural context as to why a dialogue selection was culturally appropriate or inappropriate. The feedback system did not provide explicit references to cultural dimensions but rather to the context of the specific interaction. An example of this

TABLE 4.1: Hofstede Dimensions represented in each Scene

Scene	**Hofstede Dimensions**
Classroom	Power Distance, Collectivism, Restraint
Hallway	Long-Term Orientation, Restraint
Canteen	Long-Term Orientation, Restraint
Library	Long-Term Orientation, Restraint
Office Hours	Power Distance, Collectivism

feedback is shown:

Avatar Response: 你好，我是李敏，是和你一组的不知道你今天晚上六点有空吗？ (English translation) *Hello, I'm Li Min. We are in a group together. Are you free this evening at 6 pm?*

Player Response: 应该不行。(English translation) *No, that won't work.*

Feedback: *Being so direct when declining an offer can damage long-term relationships*

The final two scenes served as test environments and the feedback system was disabled. Table 4.1 indicates which cultural dimensions were addressed in each scene.

Classroom

The classroom scene simulates a conversation with a familiar peer beginning with general small talk and also to set up a meeting time to work on a collaborative assignment. The objective of this scene is to familiarize the participant to the simulation dynamics.

Hallway

The Hallway scene simulates a conversation with an unfamiliar peer with regards to negative cultural stereotypes and other often difficult discussion topics. This scene exposes the participant to difficult conversations whilst exercising the use of formality in the Chinese language.

Canteen

The Canteen scene also simulates a conversation about stereotypes but this time with a familiar peer. This scene attempts to capture linguistic differences when conversing with a familiar versus an unfamiliar peer.

Library

The Library scene simulates a conversation about the project discussed in the Classroom scene. As the first test scene, this scene revisits dialogues exercising the previously exposed Cultural Dimensions.

Office Hours

The Office Hours scene simulates a formal conversation with a professor to get assistance with an assignment. This scene also revisits target Cultural Dimensions for the purpose of assessment.

FIGURE 4.2: In-game Library scene with Mandarin Dialogue

4.3 Cultural Dimension Design Process Case Study

In order to demonstrate the use of the previously described framework, we conducted an experiment to evaluate user performance after having been exposed to a training game developed with this framework. Prior to the experiment, the dialogue validity was attained through a multiple rater evaluation leading to an interrater reliability calculation reported in previous work [14].

The experimental design of this study was a 2 (Game Exposure) x 2 (Gender) mixed factorial design. Both factors were between-subjects factors. Language skill in Chinese was applied as a blocking variable given the growing body of research showing correlations between language skill and 3C[89][40]. The two Game Exposure factor levels were the control group and the trained group where the trained group was exposed to the full game. Given previous findings supporting differences in 3C, gender was also evaluated[29][19][7]. A randomized block design was used to assign subjects to either the control group or the test group[86].

The independent variables considered were each of the four Cultural Intelligence factors developed by Ang et al.[8], demographic information, language skill, and test group. The dependent variable was an in-game score based on the appropriateness of the subject's in-game selections. Cultural Intelligence was considered as an independent variable to serve as a way to assess the subjects' inherent cultural competence prior to exposure to the game. Given that the experiment was intended to measure performance growth from the game, it was important to capture each subject's natural 3C. The heritage and validity of Cultural Intelligence are described below.

In summary, this experiment was designed to address the following high-level questions. Does a game designed using the 3C Game Design Framework result in a more profound understanding of cultural differences? Does this understanding result in better performance, consistency, and applicability in new situations?

Hypotheses

The hypotheses of the research are described below.

H1: Subjects exposed to the CCDTS will score higher than subjects without training when controlling for Cultural Intelligence.

H2: The scores of subjects exposed to the CCDTS will have a smaller variance than the scores of subjects without training.

Game

The test group was exposed to all five scenes with the first three acting as the training scenes and the last two acting as the evaluation scenes as depicted in Figure 4.3. Using an in-game ITS, the subjects were provided feedback in the first three scenes as to the appropriateness of their in-game selections. The feedback provided was

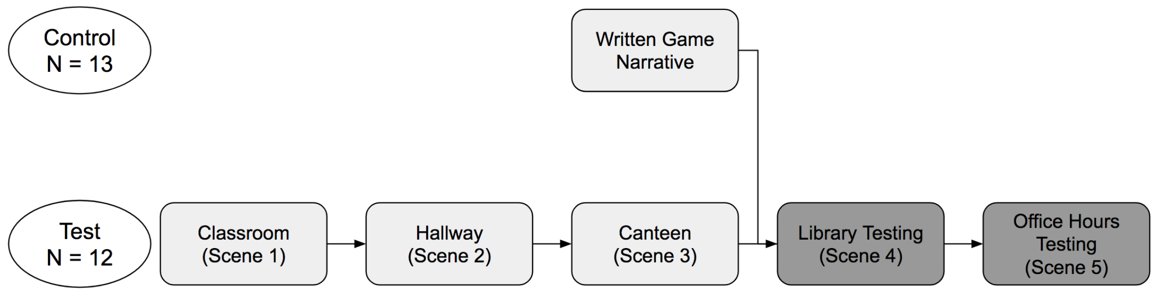


FIGURE 4.3: Depiction of Experimental Design

specific to the dialogue selection and made no reference to the cultural dimension being presented nor the concept of cultural dimensions in general.

The control group was presented with a narrative description of the first three scenes in order to provide context to the story line. The narrative description was necessary as each of the scenes built sequentially upon each other. Once the subjects had completed reading the descriptions, they were presented with the last two in-game scenes (Library and Office Hours) which served as the evaluation. Ideally, we would have had a second control group that played a similar duration game but without a cultural focus. This additional control would have helped identify whether game performance was attributable to game play in general or specifically to the CCDTS. Given our limited sample size, however, this was not feasible for this experiment.

Participants

Undergraduate students (N=25) were recruited from the Chinese language school at the University of Virginia. Students who met the appropriate linguistic and cultural knowledge were offered a $10 gift card. A requirement imposed on participants was that they did not primarily identify as culturally nor ethnically Chinese. Additionally, participants were required to have successfully completed a 2000-level Chinese course and they were also not allowed to have learned Chinese as their first language. These requirements were a limiting factor to our participant population since this demographic primarily existed in the language school.

Additionally, of the 25 participants, there were 10 males and 15 females. The participant's mean age was 20.67 and the median was 20.

Procedures

Each experimental session lasted approximately 0.75h. Each subject was first asked to sign an informed consent, fill out a pre-test questionnaire, and was then presented to the game. The pre-test questionnaire consisted of a demographic survey and the Cultural Intelligence Inventory. Each experiment was held in a research lab and the serious game was presented on a 28in. monitor where a computer mouse was the primary input device. Prior to the start of the session, each participant was given a five minute tutorial on the game dynamics and an opportunity to become familiarized with the interface. Based on the assigned test group of the subject, they were either presented the full five scene game or a game narrative and the abridged two scene game. Each subject was tested independently.

Measures

Cultural Intelligence (Independent Variable)

In addition to the test group in which a subject was assigned, we recognize that a subject's performance is also affected by inherent personality traits and natural cultural competence which needed to be accounted for. To this end, a modified version of the CQ Inventory was the selected measurement methodology given its growing popularity and recent validation efforts[8][136][83]. The modification we made to the original inventory was to duplicate the Cognitive component of the inventory whereby the duplicate version targeted knowledge of Chinese culture specifically. An example of this modification is shown:

Original: *"I know the legal and economic systems of other cultures"*

Modified: *"I know the legal and economic systems of China"*

In-Game Score (Dependent Variable)

An in-game scoring methodology was generated during the development of the CCDTS. During the evaluation scenes, each dialogue option that is presented to the subject was assigned a numeric score from one to three based on the degree to which it matched the spectrum of the cultural dimension being tested. The scores of each response were derived from a SME evaluation of the dialogue responses[14]. Once a subject completes an evaluation scene, the total points scored is divided by points available and a percentage score is reported. The purpose of this normalization is that based on the dialogue options selected each subject is presented with a different number of dialogue instances resulting in varying degrees of maximum available points. Each evaluation scene (Library and Office Hours) produced a score which were assessed independently since each scene evaluated different cultural dimensions as shown in Table 4.1.

Data Analysis

The data analysis in this experiment was performed to primarily identify performance gains of in-game score in each of the two evaluation scenes. Stepwise Multiple Linear Regressions was used to determine if an adequate model existed by which the in-game score could be predicted by the treatment group and some subset of the independent variables. Ordinal variables were binary coded.

The objective of Stepwise regression is to achieve the maximum estimation power of a model with the fewest number of independent variables. This selection can be made based on various criteria, but the Akaike Information Criterion (AIC) was used in this analysis[3]. The significance level of the resultant model and its respective factors were used to determine whether any other independent variables were significant to predicting the in-game score.

Additionally, the variance of the in-game scores by treatment group was also analyzed to assess the consistency of the training and results. An F-Test to compare variances was performed as an indicator to capture potential 'guessing' behaviors resulting in artificially higher or lower scores.

Results

Overall, our results substantiated our hypotheses as described below. The independent variables considered in the models shown in Tables III-VI are defined in Table 4.2.

TABLE 4.2: Independent Variable Descriptions

Variable	**Description**
Test Group	Participant is in Test Group
Male	Participant is Male
CQ Behavior	Behavior Component Score of CQ
CQ Motivation	Motivation Component Score of CQ
CQ Strategy	Strategy Component Score of CQ
CQ Knoweledge(China)	Knowledge Component Score of CQ

TABLE 4.3: Library Scene Stepwise Regression Model

	Dependent variable:
	Library_Score
Test Group	0.097^{**}
	(0.032)
Male	-0.093^{**}
	(0.038)
CQ Behavior	-0.043^{***}
	(0.013)
Constant	1.107^{***}
	(0.062)
Observations	15
R^2	0.576
Adjusted R^2	0.460
Residual Std. Error	0.055 (df = 11)
F Statistic	4.976^{**} (df = 3; 11)
Note:	*p<0.1; **p<0.05; ***p<0.01

Game Exposure effects (H1)

H1 stated that subjects exposed to the entire CCDTS would not score higher than the control subjects. Stepwise selection produced the model in Table 4.3 for the Library scene. This model has the treatment group as a significant factor as well as gender

TABLE 4.4: Library Scene Stepwise Regression Model w/ All CQ Factors

	Dependent variable:
	Library_Score
Test Group	0.081^{*}
	(0.037)
Male	-0.086^{*}
	(0.040)
CQ Motivation	-0.012
	(0.029)
CQ Behavior	-0.038^{**}
	(0.015)
CQ Strategy	0.047
	(0.033)
CQ Knowledge(China)	-0.043
	(0.031)
Constant	1.124^{***}
	(0.163)
Observations	15
R^2	0.684
Adjusted R^2	0.447
Residual Std. Error	0.056 (df = 8)
F Statistic	2.884^{*} (df = 6; 8)
Note:	$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

TABLE 4.5: Office Hours Scene Stepwise Regression Model

	Dependent variable:
	Office_Hours_Score
Male	-0.081^{**}
	(0.033)
Constant	0.941^{***}
	(0.019)
Observations	15
R^2	0.319
Adjusted R^2	0.267
Residual Std. Error	0.060 (df = 13)
F Statistic	6.089^{**} (df = 1; 13)
Note:	*p<0.1; **p<0.05; ***p<0.01

and the Behavior component of CQ. As a point of comparison, the model with all the CQ factors is shown in Table 4.4. H1 is rejected for the Library scene with $p < 0.05$.

Additionally, a stepwise model was created to predict the Office Hours scene score as seen in Table 4.5. Aside from gender, the stepwise procedure eliminated all the factors in this model including the treatment groups. Table 4.6 shows the regression model with all the CQ factors highlighting the lack of significance for any of the independent variables. H1 was not rejected for the Office Hours scene.

Variance of Game Exposure effects (H2)

H2 stated that the scores of subjects exposed to the entire CCDTS would exhibit a lower variance as opposed to the scores of subjects who were not. An F-test on variances shows a significant difference for the Library scene [F(6,7)=11.152, p = 0.00556). For the Office Hours scene, there is no significant difference [F(6,7)=3.1793,p=0.1559].

TABLE 4.6: Office Hours Scene Stepwise Regression Model w/ All CQ Factors

	Dependent variable:
	Office_Hours_Score
Test Group	0.005
	(0.046)
Male	−0.080
	(0.050)
CQ Motivation	−0.028
	(0.037)
CQ Behavior	0.005
	(0.020)
CQ Strategy	0.043
	(0.042)
CQ Knowledge(China)	−0.028
	(0.040)
Constant	0.978***
	(0.207)
Observations	15
R^2	0.405
Adjusted R^2	−0.041
Residual Std. Error	0.071 (df = 8)
F Statistic	0.909 (df = 6; 8)
Note:	$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

4.4 Discussion

The results from the linear regression model seen in Table 4.3 and the significant difference in variance indicate that the CCDTS developed with the framework is effective in increasing in-game score. Several studies in the area of Cross Cultural Competence have produced survey-based methods for showing long term developmental progress in 3C but this study is the first attempt to our knowledge to quantify changes in measures relating to 3C from short term training systems like the CCDTS. Further research in the area of quantitative developmental 3C models is needed to fully and justifiably correlate the in-game score presented in this research with the theoretical concept of 3C.

Several notable points were observed in our results that warrant further discussion. Specifically, it was observed that the Behavior component of CQ was significant but had a negative coefficient indicating that higher self-reported CQ Behavior correlated to a lower in-game score percentage. Intuitively it would seem that if any CQ component was significant, it would result in a positive coefficient. Ang et al. also found a negative correlation between behavioral CQ and emotional stability[8]. This finding was explained by the possibility that the calmness associated with emotional stability inhibits appropriate behavioral responses. This explanation is plausible for our finding as well given the body of research supporting the correlation between emotional intelligence and educational/occupational performance[111][103].

Though the F-test on variance showed a significant difference in variance between the control and test group, it should be noted that we were unable to show a significant difference in mean scores. The conclusion that we draw from this is that training through this game framework results in more consistent performance though this finding requires further validation with larger sample sizes.

Limitations

Given the pedagogical motivation of this experiment, we instituted a requirement that all participants have Chinese language skills but not be native or culturally Chinese. In addition to this constraint, participants needed to have adequate Chinese language abilities to understand the prompts being presented. Not only did this limit the participant search population to those individuals in the university language school, but to those participants that were in higher level classes. Future versions of this game must allow for these constraints to be relaxed in order to increase the statistical confidence of the results.

Additionally, we recognize that the measure of performance being an in-game score unique to this application does not lend itself well to being generalized to other applications. This also should be addressed through more general performance measures such as cognitive or physiological measures.

Separately, we used the pre-test CQ inventory as an indication of pre-game 3C. Ideally, a consistent pre/post test needs to be developed to show the incremental progress of tools such as the CCDTS.

In an effort to further substantiate the findings in this chapter, it would have been ideal to have had a second control group whereby the participants would have played a similar game without a focus on 3C. This would have allowed us to determine whether increases in performance were specifically attributable to the CCDTS or to general game play.

Future Research

As of this writing, this study is one of the first attempts to capture cultural dimension training and assessment in the dialogue structure of a computer-based serious game. As such, there exists a number of potential research directions worth noting.

For instance, the post-game assessment was an in-game scene designed to represent, but not mimic, the cultural dimensions presented in the training scenes of the CCDTS. An ideal approach to making the post-game assessment would be through dialogue with real individuals. Having a predetermined set of responses presumably increases the probability that an answer is correctly guessed and thus result in a higher score without actually increasing an individual's 3C.

Also, we used the Hofstede dimensions as our basis for more profound 3C, but there are a number of other cultural models such as Rasmussen et al's ARC that exist and continue to become more relevant through third party validation. These models also need to be explored within this context[115].

The testing of long-term knowledge retention is also recommended to further inform the lasting efficacy of this framework. Our assessment was conducted immediately following the conclusion of the training scenes and did not capture the long-term effects of the CCDTS.

We believe that future research in this area has the potential to give the military, businesses, and other users of this framework the means to make early assessments of 3C of their personnel. These kinds of assessments can be used to make selections for promotion, overseas assignments, or even to manage domestic leadership roles of diverse teams. Though not explicitly shown in this study, it is imperative to show whether frameworks like this are capable of imparting universal cultural knowledge and not strictly competence for the particular culture in the game.

4.5 Conclusions

Using the novel Cultural Dimension Design Process, the CCDTS was developed to implicitly train and assess an individual's knowledge of Cultural Dimensions in Chinese culture. Participants who played the entire CCDTS scored higher than those that only played the evaluative segment of the game. This finding contributes to the theoretical knowledge of how to design Cultural Serious Games in a way which targets specific training cultures and the respective backgrounds of the participants. The method described uniquely leverages existing knowledge from the field of Cultural Psychology to allow an individual to gain a deeper understanding of culture which can potentially be applied beyond the situations presented in the training. These findings have implications for the development of cultural training tools and assessments in a variety of applications to include military, political, and business domains. In educational environments, it presents a method by which educators can implicitly teach cultural values and dimensions. Leveraging Cultural Psychology theory to create scalable digital training in the CCDTS to increase cultural communication is a first step to creating an immersive and adaptive cultural game specifically targeting the implicit learning of Cultural Dimensions. While this particular effort explored the game development framework and presented a practical example of this in a Chinese cultural game, additional testing is required with other cultures and demographically varied subjects to fully validate this method.

Chapter 5

Exploring the Impact of Immersion in Cultural Games

This chapter is based on an article accepted to be published as a conference article in the 2018 International Conference on Computer-Human Interaction Research and Applications.

As the world becomes more interconnected through global business and coalition military operations, it has become a necessity for individual's to exhibit Cross Cultural Competence (3C). An individual's 3C is characterized by the Department of Defense as the "set of knowledge, skills, and affect/motivation that enable individuals to adapt effectively in cross-cultural environments[55]." This study evaluates the effectiveness of a game designed to improve 3C delivered in a Head-Mounted Device (HMD) Virtual Reality (VR) to the more traditional desktop version of the game. While the evolution of HMD VR technology has garnered much attention with the introduction of devices such as the Oculus Rift and the HTC Vive, few studies have empirically assessed the added effectiveness of HMD VR systems over traditional desktop channels in improving social-interaction skills such as 3C. This study specifically investigates the effectiveness of a 3C training game presented in both HMD VR and a traditional desktop medium in order to inform future investments of 3C training systems.

5.1 Background

Since the introduction of game-based learning, 3C games have taken many forms[48]. One of the first games, BAFA BAFA™, was designed as a moderated board game where participants would take on the role of fictional cultures and attempt to communicate and collaborate with other participants[48].

As gaming technologies improved and theories behind 3C evolved, more sophisticated computer-based games emerged[90][132][45]. These games targeted various aspects of 3C to include communication, negotiation, culture-specific knowledge, and behavioral norms.

Though these games have demonstrated varying degrees of effectiveness in improving 3C, one area that has yet to be investigated is the use of highly immersive modern VR systems to train 3C. Given the cognitive and sensory attributes associated with 3C, it would seem highly plausible that VR could be a better learning environment for 3C than the more traditional digital means previously described[137][99].

Virtual Reality

At its most abstract definition, Virtual Reality has been defined as "an artificial environment which is experienced through sensory stimuli (as sights and sounds) provided by a computer and in which one's actions partially determine what happens in the environment[102]." Given this definition, the term Virtual Reality is interchangeably used for a wide variety of technologies and experiences. Previously, Virtual

Reality described lower quality and lower specification systems using computer or projection screens to display content[50]. Compared to today's standards of interaction and immersion, these systems would hardly be considered virtual reality. Modern virtual reality systems, however, take many forms. Of these systems, two technologies have emerged as industry leaders: Cave Automatic Virtual Environments (CAVE) and Head-Mounted Device Virtual Reality(HMD VR) systems.

The CAVE system usually consists of four projection screens to serve the content using stereo projection. Within the the environment, users wear a set of tracked equipment ranging from simple 3D glasses to full body suits to capture all body motion. Generally, tracking is conducted either with inertial sensors or more commonly with infrared markers[80][81].

HMD VR systems utilize a notably different methodology to achieve immersion. HMD VR systems consist of a goggle-like headset which projects an independently rendered image for each eye which corrects for interpupillary distance. The HMD position is tracked (typically with infrared markers) in order to accurately render and project the appropriate image of the head-based gaze direction. Rapid refresh rates of the projection allow for users to naturally move their head and continuously see a real-time 360 degree stereo projection of the scene(See Fig 5.3). Due to its widespread commercial adoption and optimized content development tools, HMD VR systems have been the target of many commercial entertainment and educational research ventures[13][49].

Immersive VR DGBL Systems

Recent innovations and cost-efficiencies with immersive HMD VR have spurred a lot of interest with its use in digital game based learning (DGBL) systems specifically where sensory immersion is hypothesized to augment learning benefits. Freina et al propose that VR can add learning value in situations that "cannot be physically accessed[50]." They postulate that this limitation can be due to a number of reasons to include temporal limitations (i.e. situation occurred in a historic time period), safety limitations (i.e. hostile or emergency situations), and ethical limitations (i.e. performing a high-risk surgery by non-experts). Though studies have shown that immersive technologies such as HMD VR can increase Quality of Experience and Engagement, there is a lack of empirical evidence comprehensively characterizing the learning advantages and performances of these technologies[70].

Freeman et al conducted a comprehensive literature review of empirical studies of the use of virtual reality in mental health treatments and found large inconsistencies in the use of the term 'virtual reality' with many of the recent studies not actually referring to modern VR as previously described[49]. In the experiments that, in fact, did experiment with modern VR, results were mixed. In one instance, a research group found no difference in performance between an HMD VR and a desktop version of DGBL system designed to teach Spatial Perspective Taking to mildly intellectually impaired teenagers[50][51].

Also, multiple studies have assessed the impact of immersive VR on teaching various academic concepts. A research group found that when participants performed a musical signal flow task in HMD VR versus a traditional computer screen, no performance difference was found[130]. Makransky et al conducted an experiment comparing immersive VR an a traditional screen to teach science concepts. Using electroencephalogram (EEG) measurements, they found that though immersive VR increased sense of presence, it also resulted in poorer learning outcomes[93]. On the contrary, other studies have shown positive learning results in immersive

VR[4][141]. In the area of academic learning systems, the current research is inconclusive as to whether immersive VR results in better learning.

In the area of social and cultural serious games, we found a limited number of studies evaluating performance in immersive VR. One related area of research addresses the effects of immersive VR in treating cognitive and social disabilities. Several studies have shown promising results in the use of immersive VR to treat High-Functioning Autistic children though the results are far from conclusive and further research is recommended before any treatment recommendations are considered[12][20]. Therefore, the limited number of studies make it necessary to further investigate the learning benefits of social and cultural serious games.

Immersion and Gender

Previous studies comparing various levels of immersive game environments have shown differences in learning performance between males and females. Though no studies were found evaluating modern immersive VR systems, several studies compared traditional or lower fidelity VR systems to lower immersion systems and found that males generally performed better.

Given previous findings that females use virtual spaces for communal purposes, Coffey et al hypothesized that females would score higher in a 3D virtual environment (non HMD-VR) versus a static web environment[57][29]. However, from their experiment, they found the opposite to be true and that, in fact, males showed more improvement in the 3D virtual environment than females.

In an experiment measuring recall after being delivered a lecture in VR, Bailenson et al found that males performed significantly better than females[11]. They hypothesized that these gains could be attributable to findings reporting that men were more experienced using video-games[148].

Ausburn et al specifically investigated the gender effect in virtual learning environments and proposed a theoretical framework to explain the difference[10]. They hypothesized that these differences can be attributed to gender differences in "Technology Self-Efficacy" characterized by a set of experiences, skills, and perceptions that various studies have also shown gender differences (See Fig 5.1)[10].

Though our literature review did not find any specific studies that found similar gender differences in modern immersive VR systems, the previously mentioned studies would lead one to believe that males would also achieve better results in these systems.

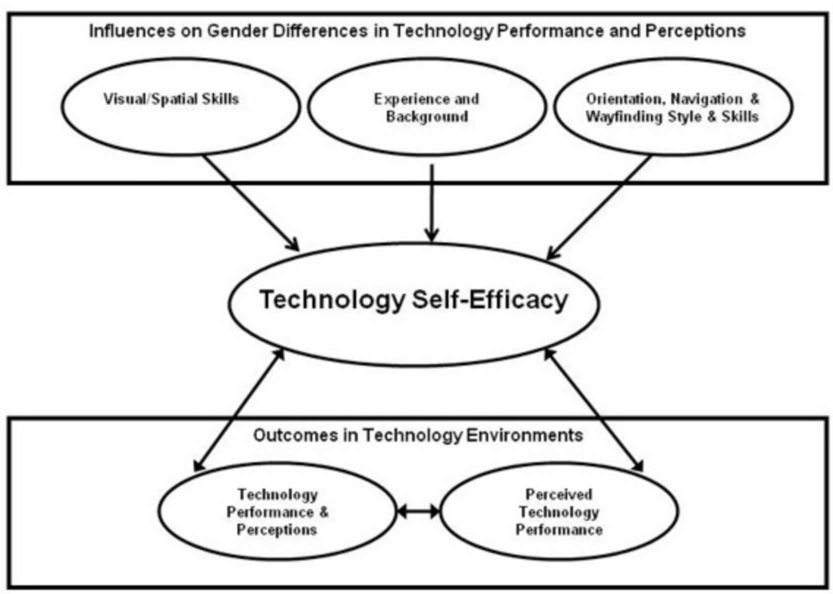


FIGURE 5.1: Ausburn's theoretical framework explaining gender differences in virtual learning environments[10].

5.2 Research Questions

The primary research question of this study is to investigate whether a 3C game played in immersive HMD VR has a significant effect on the acquisition of 3C as compared to a game played on a traditional desktop medium. Given previous works related to HMD VR, we hypothesize the following:

- **H1:** Participants exposed to the HMD VR version of the CMTS will exhibit greater improvement than those that are exposed to the desktop version.
- **H2:** Males exposed to the HMD VR version of the CMTS will exhibit greater improvement than females.

5.3 Experimental Design

Our experimental design was a 2 (Game Version) x 2 (Gender) mixed factorial design. Using a randomized-block design, participants were randomly assigned to either the traditional or VR group blocking for reported Cultural Intelligence (CQ) score. Cultural Intelligence is a self-report inventory designed to evaluate a participants 3C developed by Van Dyne et al[137]. Recent independent studies have shown the CQ model to be a promising measure of 3C and related personality traits[99][5][27]. In order to ensure an even distribution of pre-experiment 3C, participants were blocked around a predetermined CQ score threshold.

Stimuli

The stimuli used in this study was a custom developed serious game called the Joint Chinese American Cultural Simulator(JCACS) designed to teach participants Cross

Cultural Competence. The game was designed using the Cultural Simulation Design Process[6]. The game is comprised of two military training scenarios whereby the participant plays the role of a US Army Officer in charge of humanitarian efforts. In each scenario, the player is expected to work with a Chinese People's Liberation Army Officer in order to conduct joint humanitarian efforts. Players are presented with dialogue options of varying degrees of appropriateness. When a dialogue option is selected, a feedback system provides specific guidelines as to why a particular selection was or was not appropriate. In the first scenario called the Disaster Management Exercise (DME) scenario, US and Chinese forces are working together to provide disaster relief to an Ebola-stricken community in the Liberian Lowlands (see Fig 5.2). In the second scenario called the Earthquake scenario, US and Chinese forces are providing aid and assistance to injured civilians after a deadly earthquake in Nsunga, Tanzania. Both versions of the game had identical visual and dialogue content.

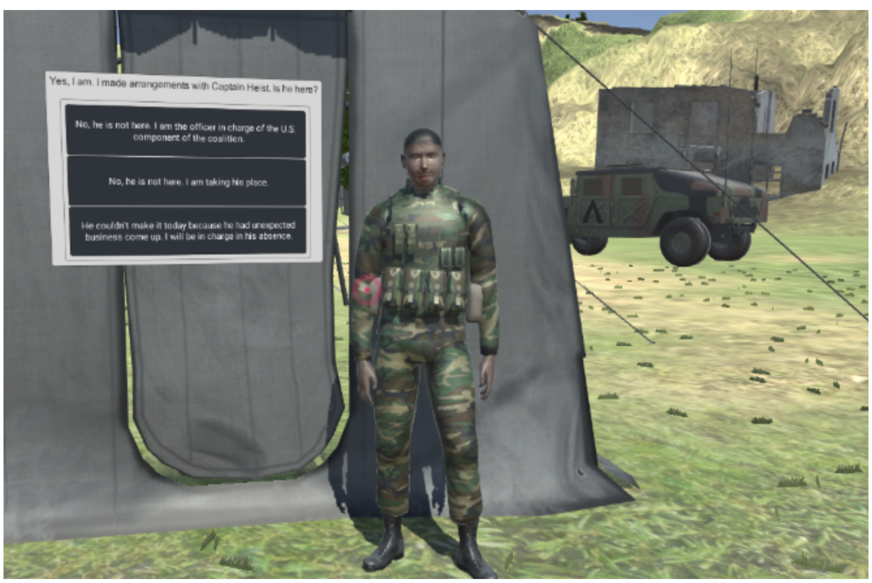


FIGURE 5.2: The player is interacting with the PLA Officer-in-Charge in the DME Scenario of JCACS

Desktop Variant

Though both versions of the game have identical visual and dialogue content, each game version has notable differences in User Interface. The Desktop version has a locked camera preventing the user from changing what was visible on the screen. The dialogue system is locked to the bottom of the viewing area and the primary mode of user interaction is by clicking on the preferred dialogue option.

HMD VR Variant

The HMD VR version of the game has a notably different user experience and User Interface. This version was developed specifically for the Oculus Rift HMD VR system(see Fig 5.3). First, the camera motion is tied to the motion-tracking system of the Oculus Rift. This allows the participant to see the entire scene around the player. Second, the dialogue system is presented as a floating dialogue box as scene in Fig 5.2. Dialogue selections are made through prolonged gaze and simultaneous controller selection on the preferred option.

Participants

Twenty-one participants (Male=12, Female=9) were recruited for the experiment. The distribution of participants was 11 Reserve Officer Training Corp (ROTC) cadets and 10 non-military affiliated participants. The participants' mean age was 20.58 years (Median = 21, SD = 1.06). Participants were equally split into the HMD VR group (n=11) and the Desktop group (n=10). For completing the experiment, each participant was compensated with a $20 gift card.

FIGURE 5.3: A participant playing the HMD VR version of the game

Procedures

Students were individually assessed. Participants were first e-mailed an online survey intended to capture basic demographic information as well as their CQ score. This information was used to ensure equal representation in the test and control group while blocking for the relevant factors. When participants arrived for the experiment, they were first administered the pre-test. Following the pre-test, participants engaged in a tutorial specific to the game version they were playing. Following the tutorial, participants played both the DME and Earthquake scenario of the game. Once the game was complete, the participants were administered the post-test as well as a feedback survey to gauge user experience. The entire testing process for a single participant was approximately one hour. See Fig 5.4.

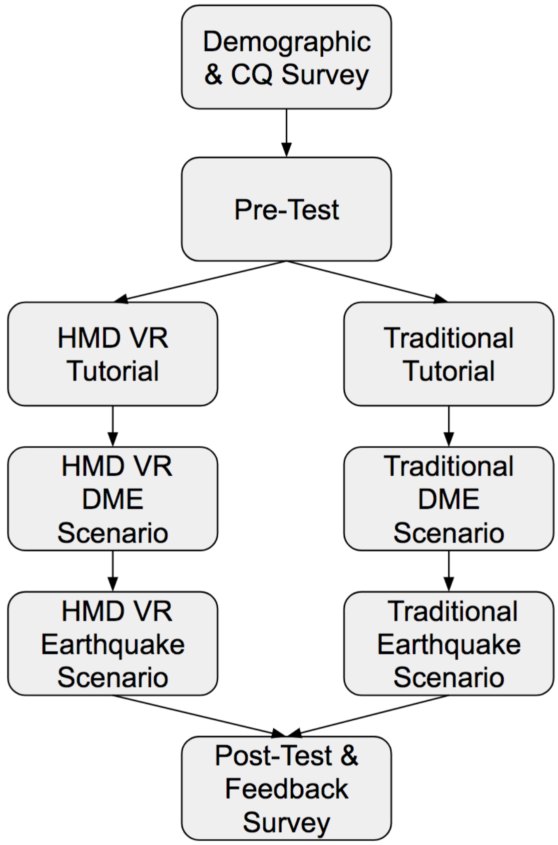


FIGURE 5.4: Experiment Flow

Measures

The pre-test and post-test used to determine performance improvement were conducted as real-time dialogue situational judgment test (SJT). SJTs have often been use to evaluate 3C though previous versions have typically been written exams with multiple choice responses[90][144]. Though traditional SJTs have been effective in measuring cognitive and metacognitive thought processes, they do not capture any performance with respect to culturally appropriate communication and behaviors. As such, we adapted the SJT to be a role-playing dialogue with a Chinese actor hereby referred to as the role-playing SJT (rSJT). The rSJT is initiated by presenting the participant with a scenario dossier which describes the role the player must play as well as the objectives that player is trying to achieve through the dialogue(See Appendix D and Appendix C). Once the participant has read through the dossier, the actor initiates the conversation with a scripted initial response and at which point the conversation becomes unscripted and free-flowing.

The rSJTs are audio visually recorded and these recordings are then rated by independent Chinese cultural experts for cultural proficiency. The experts score the rSJTs with a previously developed rubric which then produces a numeric score for each rSJT(See Appendix E). In order to ensure the reliability of the ratings, a Cohen's Kappa is calculated on the ratings[46].

Data

Collection

Between the demographic survey and the CQ inventory, a variety of data was collected on each participant. The demographic survey consisted of two categories of items characterize by 1) Personal Information and 2) International Exposure for a total of 12 unique survey items. The previously described CQ inventory is composed of 37 items subsetted into 4 main factors and 11 sub-factors (see Table 5.1) [137]. The score of each main factor or sub-factor is calculated as the average of each of the items within the main factor or the sub-factor.

TABLE 5.1: Cultural Intelligence Sub-Dimension Model

Main Dimension	**Sub-Dimension**
Cognitive	Cultural General Knowledge Context Specific Knowledge
Motivation	Intrinsic Motivation Extrinsic Motivation Self Efficacy to Adjust
Metacognitive	Planning Awareness Checking
Behavior	Verbal Behavior Non-Verbal Behavior Speech Acts

Analysis Methods

Multiple Linear Regression was the primary method used to identify any performance increases with respect to the pre/post test and the independent factors collected in the experiment. Initial factor sets were based upon factors found to be relevant in previous studies as well as those that theoretically showed the most promise. In order to identify the most significant models, AIC-based Stepwise regression was utilized to determine the most significant models[3].

5.4 Results

The results are organized by the hypotheses previously described.

H1 Analysis Results

Participants exposed to the HMD VR version of the CMTS will exhibit greater improvement than those that are exposed to the desktop version.

To examine the hypothesis H1, a Welch's t-test was performed between the HMD VR group and the desktop group with respect to the performance improvement. This t-test realized an insignificant difference (t=0.47931, df=15,521, p=0.6384).

Additionally a multiple linear regression model was generated with the independent behaviors consisting of the Metacognitive and Behavior subdimensions of CQ as well as various demographic factors as shown in Table 5.2. The Metacognitive CQ component "reflects the mental capability to acquire and evaluate cultural knowledge.[137]" The Behavior CQ component "reflects the capability to flex behaviors to fit different cultural contexts.[137]" The Metacognitive and Behavior CQ components were considered in the model for their specific relevance to skills that we hypothesized would impact the pre-test and post-test scores.

This model showed a negative significant correlation for those participants in the HMD VR group. An AIC stepwise regression of that model also resulted in a similar negative significant correlation as seen in Table 5.3.

H2 Analysis Results

Males exposed to the HMD VR version of the CMTS will exhibit greater improvement than females.

With respect to H2, we explored the significance of an interaction effect between gender and VR. The linear regression model shown in Table 5.2 showed a significant interaction effect as seen in the model as "VR:genderMale." Additionally, the AIC stepwise regression resulted in a similar effect though the "VR:genderMale" interaction effect was removed as seen in Table 5.3.

TABLE 5.2: Initial Regression Model of Performance Improvement

	Dependent variable:
	Improvement
VR	-0.356^{*} (0.187)
meta_Planning	-0.117^{*} (0.064)
meta_Awareness	0.035 (0.103)
meta_Checking	0.364^{***} (0.105)
beh_Verbal	0.062 (0.084)
beh_Non_Verbal	0.097 (0.054)
beh_Speech	-0.202^{*} (0.095)
genderMale	-0.073 (0.184)
ROTC	0.169 (0.140)
VR:genderMale	0.484^{**} (0.180)
VR:ROTC	0.169 (0.184)
Constant	-1.531^{**} (0.574)
Observations	21
Adjusted R^2	0.595
RSE	0.173 (df = 9)
F Statistic	3.672^{**} (df = 11; 9)
Note:	$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

5.5 Discussion

The findings are discussed by the previously stated hypotheses.

The first hypothesis investigates whether the HMD VR variant of a game results in more learning performance as compared to the traditional desktop variant. We

TABLE 5.3: AIC Stepwise Regression Model of Performance Improvement

	Dependent variable:
	Improvement
VR	-0.277^{*} (0.148)
meta_Planning	-0.082 (0.047)
meta_Checking	0.356^{***} (0.075)
beh_Non_verbal	0.107^{*} (0.050)
beh_Speech	-0.170^{**} (0.062)
GenderMale	-0.091 (0.164)
ROTC	0.292^{***} (0.084)
VR:genderMale	0.466^{**} (0.166)
Constant	-1.407^{**} (0.503)
Observations	21
Adjusted R^2	0.647
RSE	0.161 (df = 12)
F Statistic	5.575^{***} (df = 8; 12)
Note:	$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

found that, generally, the two variants did not perform differently when other factors were not considered. Additionally, when the test group was considered along with the demographic factors, it was found that generally the HMD VR group performed slightly worse that the control group as shown in Tables 5.2 and Table 5.3. This finding parallels the results of studies in other domains that found no difference or a negative impact for HMD VR[130][93]. It is possible that the perceptual realism added through HMD VR is in fact distracting. Through EEG measurements, Makransky et al found that HMD VR increased cognitive load as compared to desktop versions of the same biology lab learning system. They explained that this increased load may overload participants and thus "result in less opportunity to build learning outcomes.[93]"

In analyzing the results for the second hypothesis, we discovered a more nuanced explanation to the conclusion drawn from the first hypothesis. The second order effect between participants who were male and were in the HMD VR test group was found to be highly significant. The positive coefficient of this interaction would lead one to conclude that HMD VR does, in fact, have a positive effect on learning outcomes for males. This extends the findings of previous studies that have also found correlation between higher immersion digital learning systems and gender[29][10]. Specifically, we conclude that this correlation between immersion and gender can be extended through the immersive experience of HMD VR. Causality of this finding continues to be an open question that requires further investigation though the theoretical construct of technological self efficacy is supported by our finding has well as other independent findings[10].

5.6 Conclusion

This study has further advanced our understanding of the impacts of using highly immersive virtual environments in social and cultural DGBL systems as well as the specific factors that explain these impacts. We found that generally more immersive systems(e.g. HMD VR) do not necessarily result in better learning outcomes in a cultural DGBL system. However, the strong interaction found between HMD VR and male participants also indicates that demographic factors such as gender should be considered when developing cultural DGBL systems and, perhaps, all types of DGBL systems. This study motivates further research with highly immersive DGBL systems in other domains in order to draw more concrete conclusions about general immersive systems beyond those drawn in this study.

Chapter 6

Measuring Performance with Eye Gaze in Cultural Games

This chapter is based on an article that has been accepted to be published as a conference article in the 2018 Winter Simulation Conference.

The recent and rapid growth of the computer gaming industry has also served as a catalyst for the growth of an $8.1B industry around computer game-based pedagogical tools also known as digital game-based learning (DGBL) systems [1]. DGBLs have gained popularity across a wide range of applications for a number of reasons but primarily due to their repeatability and cost-effective scalability. Despite becoming a critical component to the educational economy, there has yet to be discovered an effective and cross-domain method to measure performance within DGBLs.

Initial research in this area focused on the potential learning benefits of commercial off-the-shelf entertainment games, but recently many industries and researchers have transitioned to the development and evaluation of games specifically designed for learning [30, 128]. A number of studies have aggregated and objectively evaluated games ability to meet specific learning objectives [30]. Common to many of these evaluation methods is the use of explicit user-response scores which may be susceptible to guessing thus not accurately reflecting a player's knowledge and performance. Eye-gaze technology used in tandem with DGBLs provides a novel path to measure player performance.

Eye-tracking has been successfully used in activity recognition, affect detection, engagement, cognitive load assessment, and general learning [22, 32, 72, 139]. Modern eye-tracking technologies allows for sophisticated passive process tracking with minimal cognitive load of more traditional process tracking methods such as think-aloud protocols [92]. Presupposing the validity of the Eye-Mind assumption [79], the high frequency spatial and temporal information available through eye-tracking presents an avenue by which one can infer both conscious and subconscious cognitive processes that would be too ephemeral for active process tracking methods. The general Eye-Mind assumption asserts that the object of fixation is the focus of attention and cognitive processing. Widespread research has supported the Eye-Mind assumption though a number of studies have also presented situations where it was shown to be questionable or insufficient [85] especially in situations when the visual scene is unrelated to the intended cognitive activity.

This study contributes to the growing body of research investigating the eye-tracking behaviors as they relate to performance within DGBLs. Specifically, we aim to determine whether a player's eye-tracking behavior correlates to their performance in order to substantiate a novel eye-tracking based method for evaluating player performance. During this study, we evaluate the eye-tracking data of players in a DGBL designed to evaluate participants in Chinese Cross Cultural Competence.

6.1 Background

Despite the advances in the operationalization of DGBLs, research scrutinizing the methods of measuring user performance has not demonstrated the same level of growth. A majority of DGBLs leverage post-game examinations of learning objectives in the form of pre/post testing [15]. The actual implementation of the pre/post tests span a number of methods to include evaluating player's knowledge through questionnaires/surveys[29, 44]. A common derivation of this method is the Situational Judgment Tests (SJT) designed to use the learned knowledge in practical applications [90]. These methods have been deployed on a wide scale primarily for their ease of implementation though they suffer from several notable comings. Specifically, they fail to consider whether the pre-test influenced the post-test results as seen in the case of the Second China game experiment where the pre/post tests were the same test with the questions reordered [29].

Psycho-physiological data as a means of assessing learning performance within DGBLs is a rapidly growing research area. Specifically, the use of eye-tracking has gained significant traction due to the advances in both eye-tracking technology as well as DGBL development tools. Companies such as Tobii, Pupil Labs, and SensoMotoric Instruments have significantly reduced the cost and increased the resolution of eye-tracking devices while also offering a variety of collection devices for diverse collection situations. Additionally, the proliferation of popular gaming stacks to include Unity3d and Unreal Engine have made it possible for smaller research efforts to develop high quality scalable game environments for minimal costs.

With this proliferation, various applications of eye-tracking in DGBLs have been explored to determine game effectiveness. One specific research area targets the measurement of player engagement and flow through the interpretation of affect-based metrics [37, 72, 120]. A limitation, however, of affect-based measures such as pupil dilation is that they are generally one dimensional and exhibit a variable lag therefore providing limited insight into complex cognitive processes associated with learning [142].

In order to better investigate user performance, researchers have used fixation location and gaze paths to examine performance in DGBLs. Specifically, researchers have conducted experiments to identify differences in visual behavior based on skill level. Kickmeier-Rust et al found significantly different gaze path and interaction strategies between high/low performance levels of teenagers playing a game designed to teach European geography [84]. In a study examining adolescent behaviors in attention enhancement therapy, Pascual et al found that higher performance children exhibited lower fixation densities. Additionally, they were able to use ensemble machine learning methods on a combination of the gaze data and user interaction features to create a prediction model [54].

More recently, researchers have begun to investigate the use of eye-tracking to understand the process and rate of learning through gaze metrics [78, 145]. Józsa et al studied the possibility of assessing a learning curve through trends in the fixation duration and total visit duration between successive stimuli. The results, however, were mixed due to the variability of the stimuli [78].

The overall purpose of this research effort is to introduce and evaluate the efficacy of a novel eye-tracking measure of performance in social-interaction DGBL systems. In doing so, we contribute to the rapidly growing need for effective methods of measuring performance in DGBL systems.

6.2 Eye-Tracking Measure of Performance (ETMP)

Social Interaction-based serious games often have players interact with computer-driven avatars through finite dialogue trees. When a conversation commences, players are expected to sequentially select one of several dialogue options in order to progress through a conversation. In some cases, the dialogue options are either appropriate or inappropriate, but often, dialogue options have varying degrees of appropriateness scores and as such can have varying impacts on the rapport that players can build with the non-player characters. In the latter case, there is an opportunity to leverage the range of appropriateness with visual attention in order to produce a weighted score of appropriateness as described below:

$$Player_{Score} = \sum_{i=1}^{x} \sum_{j=1}^{y} s_{ij} f_{ij}$$

where s_{ij} is the score of response option j in question i, f_{ij} is the fixation proportion on response option j relative to the other dialogue options in question i, y is the number of dialogue options in a conversation instance, and x is the total number of conversation instances in the conversation. By using fixation proportions instead of absolute fixation durations, the ETMP minimizes the impact of variable reading speeds of the participants. Research has shown a gaze bias in which people fixate on options with higher ratings of subjective preference [56, 124]. Assuming the validity of the previously mentioned Eye-Mind Theory [79] and gaze bias findings, this formulation would conceptually evaluate both the appropriateness of a response as well as consider the mental schema the player uses to determine his or her response. It is important to note that this methodology only captures the overt movements of visual attention as measured through eye-tracking and makes no assumptions of

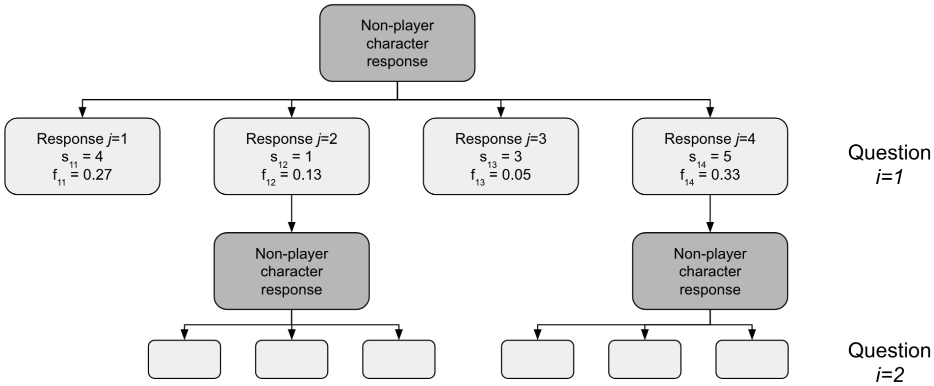


FIGURE 6.1: A graphical ETMP depiction for a single player response

the covert movements of attention that would not be observable through eye movements [38]. The gaze-based weighting of performance in this formulation results in higher scores for those players that spend more visual attention on the appropriate answers and lower scores for participants that dwell on less appropriate answers.

6.3 Research Questions

In order to investigate the feasibility and viability of a gaze-based measure of performance, this study examines how users distribute their visual attention across multiple dialogue options to progress conversations with avatars within a serious game designed to evaluate Cross Cultural Competence in Chinese Culture. The specific research questions and related hypotheses are as follows:

RQ-1) How do users distribute their visual attention across dialogue options in social interaction serious games? Do they focus their attention on preferred options and less so on rejected options?

- H_{1a}: The game participants will spend significantly more time fixating on selected answer choices than non-selected choices
- H_{1b}: The higher-scoring participants will spend significantly more time fixating on higher score responses
- H_{1c}: The advanced domain knowledge or metacognitive-ability participants will spend significantly more time fixating on higher-score responses

The intent of RQ-1 is to better understand the visual behavior of participants in the CCDTS environment. Specifically, we investigate whether the visual behavior in the CCDTS is comparable to that of other research findings in both visual behavior in multiple choice environments and other DGBL systems.

RQ-2) Does the Eye-Tracking Measure of Performance correlate to other measures of performance? Does higher domain knowledge, gaze bias, or higher traditional game scoring correlate to higher scores in the ETMP?

- H_{2a}: The game participants with advanced domain knowledge or metacognitive skill will score significantly higher in the ETMP than those participants with lower domain knowledge.
- H_{2b}: The Participants who scored higher in the traditional game scoring will also score significantly higher in the ETMP.
- H_{2c}: The participants who exhibited a gaze bias on their selected answer will score significantly higher in the ETMP.
- H_{2d}: The participants who exhibited a gaze bias on correct answers will score significantly higher in the ETMP.

The intent of RQ-2 is to provide initial validation to the ETMP through correlation to other indicators of performance whether they be in-game measurements or independent measures of Cross-Cultural Competence(3C). An individual's 3C continues to be an active area of research without a clear consensus on a 3C theoretical construct [25]. Though we discovered numerous 3C constructs and inventories in our review of the literature, we chose the Cultural Intelligence(CQ) inventory to assess an individual's 3C given promising results in recent independent validation efforts [99, 136].

6.4 Quasi-Experimental Design

In an effort to operationalize and validate the ETMP, we conducted an experiment with a DGBL system. The following section describes the quasi-experimental design, setup and analysis approach of the experiment[31].

6.4.1 Participants and Design

25 participants (40% male; M_{age}=20.67; SD_{age}=1.80) were recruited from the University of Virginia to participate in the experiment. All of the participants recruited were students in the Chinese Language Department. Each student was administered the Cultural Intelligence (CQ) Inventory to gauge self-reported 3C. 6 of the 25 participated in the study exceeded our 25% eye-tracking data loss threshold and thus had to be excluded from the analysis of the results resulting in a sample size of 19. Despite instructions to maintain visual focus in the viewing area of the computer

screen, several participants were observed consistently looking outside the limitations of the eye-tracker resulting in larger than expected track losses. Future experiments may warrant head-mounted devices in order to maintain tracking throughout the experiment.

6.4.2 Materials and Measures

The stimulus in this experiment is a custom developed first-person serious game placed in a Chinese university called the Chinese Cultural Dimension Training System (CCDTS) [14]. In the game, the user is expected to collaborate with familiar and unfamiliar Chinese students and professors while working on a group project. As the story continues, users engage in various conversations concerning politics, cultural habits, and stereotypes concluding with the player coordinating meeting times with a fellow student and visiting office hours to clarify questions with the professor. The game begins with an initial in-game training segment which introduces the storyline, allows the user to become accustomed to the game controls, and provides specific instruction to Chinese culture. It is then followed by two evaluation scenes called *Library* and *Office Hours*. In order to adapt the CCDTS to collect eye-tracking data, the bounding areas of each dialogue response had to be enlarged and had preset as Areas-of-Interest within the Tobii Pro Studio eye-tracking software.

Scene Descriptions

In the *Library* scene, the player must engage with a fellow student about the group project details and coordinate meeting times. In the *Office Hours* scene, the player engages with a professor during Office Hours to get formal feedback and recommendations about their group project.

In-Game Scoring

During the development of the game, it was necessary to score and validate the cultural appropriateness of specific dialogue responses by means of interrater reliability. For this purpose Chinese cultural experts were recruited from the Chinese language department. The raters were asked to rate each response on a Likert scale of 1 to 5 with 1 being the least appropriate and 5 being the most appropriate. The ratings determined the explicit scores associated with each response. This score was used to measure the player's performance as this experiment was not designed for a pre-test to target learning gains. The CQ inventory administered pre-stimulus, however, allowed us to investigate any pre-game cultural competency. A more detailed discussion of the scoring methodology is described by Bayer et al[14].

Given this rating system, a response with a score of either 4 or 5 was considered to be an appropriate score. Additionally, if an interaction did not have a response scored as a 4 or 5, only the highest scored response in that interaction was considered as a 'appropriate score' in determining the total fixation durations.

6.4.3 Apparatus

The stimulus was presented on a 33-in screen with a pixel resolution of 2560 by 1440. The CCDTS was played through an executable Unity file and eye-tracking data was recorded using Tobii Pro Studio V1.1. During testing, participants were seated approximately 65 cm away from the screen. The vertical placement of the screen was adjusted such that the center of the screen was at eye level of each participant. Eye-tracking data was collected using the Tobii X3 120, a screen-mounted collection system. The system was calibrated to each participant using the Tobii Pro Studio

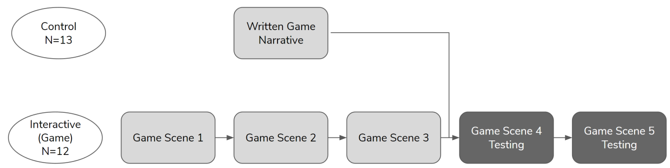


FIGURE 6.2: Experimental Design

animated 8 point calibration system. Calibration accuracy was recorded to be within 0.6of visual angle for both axes of all participants.

6.4.4 Procedure

Students were individually tested. Initially they were administered a demographic survey followed by the CQ Inventory. The CQ Inventory was used as a means to assess a players pre-game cross-cultural competence as a possible predictor of the ETMP. They were then calibrated to the eye-tracking apparatus within the tolerance of the eye-tracker and the CCDTS was initiated. The first scene in the game was a basic tutorial to teach participants how to navigate the interface and play the game. Once the tutorial was completed, the participant immediately entered the real game environment in which they interacted with non-player characters using the multiple choice style dialogue system common to modern video games. This test flow is depicted in Fig 6.2. Following the completion of the game, a short after-action interview was conducted with the participant to attain any feedback about the game experience. The information collected from these interviews are not reported in this chapter.

6.4.5 Data Collection and Analysis

Prior to analyzing the data, the fixation durations on specific areas of interest (AOI) had to be processed from the collected gaze data. Since each interaction presented a series of responses for the participant to evaluate, the container around each of the responses was considered an AOI. All fixations within the container were aggregated to determine the total fixation duration in any particular response AOI.

Since each participant was exposed to different interactions based on their response choices, few participants saw the same sequence or set of interactions. Additionally, each interaction had varying numbers of response choices meaning that the total fixation times for any particular interaction was expected to vary significantly since participants would presumably require more time to consider more responses. In order to account for the variability between subjects, *percentage of fixation duration* was calculated for each interaction [135].

The primary methods used to test each of the hypotheses were paired t-tests and multiple linear regression.

6.5 Results

This section organizes the results by each hypothesis.

H_{1a}: Participants will spend significantly more time fixating on selected answer choices than non-selected choices

To examine this hypothesis, four Bonferroni-adjusted post-hoc paired t-tests were conducted on the percentage of fixation duration. The first paired t-test conducted compared the percentage of fixation duration on the selected options and the combined fixation duration percentage of non chosen options. The second paired t-test conducted compared the percentage of fixation duration on the selected option and the average percentage of non chosen options. This set of paired t-tests was individually conducted on the data of each scene.

The paired t-test for Scene 4 comparing the percentage of fixation duration on selected options(M=0.2246, SD=0.06645) and the combined duration percentage(M=0.3117, SD=0.1055) was significantly different(t=-4.2552, df=18, p=0.00048). The paired t-test for Scene 4 comparing the percentage of fixation duration on selected options(M=0.2426, SD=0.06645) and the average duration percentage(M=0.1232, SD=0.084) was significantly different(t=7.4842, df=18, p=6.247e-07).

The paired t-test for Scene 5 comparing the percentage of fixation duration on selected options(M=0.1985, SD=0.054) and the combined duration percentage(M=0.4224, SD=0.108) was significantly different(t=-7.613, df=18, p=4.929e-07). The paired t-test for Scene 5 comparing the percentage of fixation duration on selected options(M=0.1985, SD=0.054) and the average duration percentage(M=0.12315, SD=0.0298) was significantly different(t=4.9635, df=18, p=0.0001005).

These results indicate that participants generally spent more time fixating on selected options than any other option.

H_{1b}: Higher scoring participants will spend significantly more time fixating on higher-score responses

During the development of the game, it was necessary to score and validate the cultural appropriateness of specific dialogue responses by means of interrater reliability. The raters were asked to rate each response on a Likert scale of 1 to 5 with 1 being the least appropriate and 5 being the most appropriate. The ratings determined the scores associated with each response. Given this rating system, a response with a score of either 4 or 5 was considered to be a high score. Additionally, if an interaction did not have a response scored as a 4 or 5, only the highest scored response in that interaction was considered as a 'high score' in determining the total fixation durations.

H_{1b} was analyzed through linear regression employing the average Total Fixation on high scores as the dependent variable and the game score as the independent variable. The average Total Fixation on high scores was calculated by summing the fixations on high scores for a participant and dividing by the number of interactions the participant was exposed to throughout the game play labelled in the model as "Correct Fix S4" and "Correct Fix S5" in Table 6.1. This analysis was performed separately on Scene 4 and Scene 5.

The Scene 4 model did not show the Game Score as being a significant predictor for the average Total Fixation on high scores. Scene 5 did show the Game Score as being a significant predictor for the average Total Fixation on high scores. The positive coefficient in this model indicates that higher game scores in Scene 5 were correlated to higher fixation duration percentages on high score responses.

H_{1c}: The participants with advanced domain knowledge or metacognitive ability will spend significantly more time fixating on higher-score responses

For the purposes of this hypothesis, higher domain knowledge and metacognitive ability are characterized by a participant's CQ score. Specifically, domain knowledge is captured by the Cognitive factor of CQ and metacognitive ability is captured by

TABLE 6.1: Regression Analysis for H_{1b}

	Dependent variable:	
	Correct_Fix_S4	Correct_Fix_S5
Game_Score_S4	0.006 (0.007)	
Game_Score_S5		0.008^{**} (0.004)
Constant	0.202 (0.163)	0.166^{*} (0.081)
Observations	19	19
Adjusted R^2	−0.015	0.186
Residual Std. Error (df = 17)	0.101	0.063
F Statistic (df = 1; 17)	0.738	5.118^{**}
Note:		*p<0.1; **p<0.05; ***p<0.01

the Metacognitive factor of CQ. Based on a regression analysis where the dependent variable was the total Average fixation duration percentage on correct answers, both CQ factors were not shown to be significant as shown in Table 6.2. Additionally, we analyzed the data to determine if the fixation duration percentage on incorrect responses was correlated to the CQ factors, but this also did not produce statistically significant factors as shown in Table 6.3.

TABLE 6.2: Regression Results for H_{1c} for Correct Fixation Duration Percentage

	Dependent variable:	
	Correct_Fix_S4	Correct_Fix_S5
CQ_Metacognitive	0.006 (0.035)	−0.010 (0.024)
CQ_Cognitive	−0.018 (0.032)	0.026 (0.022)
Constant	0.400^{*} (0.201)	0.266^{*} (0.136)
Observations	19	19
Adjusted R^2	−0.104	−0.032
Residual Std. Error (df = 16)	0.106	0.071
F Statistic (df = 2; 16)	0.149	0.718
Note:		*p<0.1; **p<0.05; ***p<0.01

TABLE 6.3: Regression Results for H_{1c} for Incorrect Fixation Duration Percentage

	Dependent variable:	
	Incorrect_Fix_S4	Incorrect_Fix_S5
CQ_Metacognitive	−0.014 (0.028)	−0.022 (0.023)
CQ_Cognitive	0.002 (0.026)	−0.001 (0.021)
Constant	0.262 (0.160)	0.405^{***} (0.134)
Observations	19	19
Adjusted R^2	−0.106	−0.048
Residual Std. Error (df = 16)	0.084	0.070
F Statistic (df = 2; 16)	0.134	0.585
Note:		*p<0.1; **p<0.05; ***p<0.01

H_{2a}: The participants with advanced domain knowledge or metacognitive skill will score significantly higher on the ETMP than those participants with lower domain knowledge.

Again, domain knowledge and metacognitive skill were characterized by the specific CQ components described in H_{1c}. Not surprisingly, the Metacognitive and Cognitive component of CQ were not significant factors in both the Scene 4 and Scene 5 regression analysis as shown in Table 6.4 and Table 6.5. When an Aikeke Information Criterion (AIC) based Stepwise regression was conducted, the CQ factors were eliminated[2].

H_{2b}: The participants who scored higher in the traditional game scoring will also score significantly higher on the ETMP.

As hypothesized, participants' game scores in traditional scoring as described as "Game Score S4"(t=2.884, df=12, p=0.01374) and "Game Score S5"(t=8.832, df=12, p=1.35e-06) in Table 6.4 and Table 6.5 were positively correlated with the ETMP of both Scenes 4 and 5. The AIC-based Stepwise Regression also preserved these factors.

H_{2c}: The participants who exhibited a gaze bias on their selected answers will score significantly higher on the ETMP.

Participants who exhibited higher fixation duration percentages on their selected answers scored higher on the ETMP. The factor "Avg answer fixation S4"(t=2.884, df=12, p=0.01374) as shown in Table 6.4 represented the fixation duration percentage on selected answers for Scene 4. The factor "Avg answer fixation S5"(t=3.048, df=12, p=0.01034) as shown in Table 6.5 represented the fixation duration percentage on selected answers for Scene 5.

Additionally, the fixation duration percentage on the non-selected answers was also analyzed in the same model as described above. As scene in Table 6.4 and Table 6.5, the variables "Avg non-answer fixation S4"(t=3.247, df=12, p=0.00699) and the "Avg non-answer fixation S5"(t=4.291, df=12, p=0.00105) were also fount to be significant.

H_{2d}: The participants who exhibited a gaze bias on correct answers will score significantly higher on the ETMP.

Participants who exhibited high fixation duration percentages on correct answers as defined in the results of H_{1b} also resulted in higher ETMP scores as characterized by "Correct Fix S4" (t=1.846, df=12, p=0.08967) and "Correct Fix S5"(t=2.447, df=12, p=0.03076) in Tables 6.4 and 6.5.

TABLE 6.4: Regression Analysis for Scene 4 ETMP

	Dependent variable:	
	ETMP Scene 4	
	Original	Stepwise-AIC
CQ_Metacognitive	−0.217 (0.345)	
CQ_Cognitive	0.005 (0.327)	
Avg_answer_fixation_S4	10.043* (5.525)	10.406* (5.062)
Avg_non_answer_fixation_S4	14.869*** (4.579)	15.202*** (4.273)
Correct_Fix_S4	9.612* (5.207)	9.208* (4.868)
Game_Score_S4	0.252** (0.088)	0.248*** (0.082)
Constant	−4.158 (2.630)	−5.283*** (1.674)
Observations	19	19
Adjusted R^2	0.904	0.914
Residual Std. Error	1.024 (df = 12)	0.966 (df = 14)
F Statistic	29.234*** (df = 6; 12)	49.098*** (df = 4; 14)
Note:		*p<0.1; **p<0.05; ***p<0.01

TABLE 6.5: Regression Analysis for Scene 5 ETMP

	Dependent variable:	
	ETMP Scene 5	
	Original	Stepwise AIC
CQ_Metacognitive	−0.192 (0.178)	
CQ_Cognitive	0.179 (0.167)	
Avg_answer_fixation_S5	12.597** (4.133)	11.712** (4.014)
Avg_non_answer_fixation_S5	8.743*** (2.037)	8.699*** (1.937)
Correct_Fix_S5	9.816** (4.011)	10.817** (3.831)
Game_Score_S5	0.316*** (0.036)	0.315*** (0.034)
Constant	−5.280*** (1.252)	−5.547*** (0.823)
Observations	19	19
Adjusted R^2	0.962	0.963
Residual Std. Error	0.507 (df = 12)	0.502 (df = 14)
F Statistic	77.420*** (df = 6; 12)	118.384*** (df = 4; 14)
Note:		*p<0.1; **p<0.05; ***p<0.01

6.6 Discussion

The findings from this study are organized by the research questions previously described.

The first research question investigates general visual behaviors with respect to social interaction serious games like the CCDTS. We found that participants did, in fact, fixate more on their selections rather than the non-selected choices. This is consistent with various other studies that found longer fixation durations on selected options in multiple-choice based evaluation systems [135] [62]. It is important to highlight, however, the large standard deviations of these results which may be an artifact of the varying numbers of responses for each set of dialogue options. This warrants future experiments with consistent quantities of options per interaction to investigate this hypothesis.

Additionally, we found a positive correlation between higher scoring participants and fixation duration on the more appropriate options in the Office Hours scene but not the Library scene. An explanation of this may be that the the Library scene was played prior to the Office Hours scene and, therefore, the user was visually calibrating his behavior resulting in more varied visual patterns. Assuming the Eye-Mind Theory, we interpret this result as being that higher scoring participants spent more time considering more appropriate answer options [79] and less time considering less appropriate responses. That being the case, we postulate that concentrated visual and mental focus on appropriate responses rather than less appropriate responses more concretely indicates higher cultural competency. This then sets the foundation for the validation of the ETMP.

Our analysis did not show significant correlation between our measures of domain knowledge or metacognition represented by Cultural Intelligence and fixation on appropriate answers. Though this measure has shown promise in various independent studies, the validation of this measure continues to be an active area of research and this finding motivates future experiments with other 3C constructs that have also shown potential[25, 99].

With respect to the second research question exploring the validity of the ETMP as a measure of game performance, we found significant correlation between the ETMP and the other performance-related factors considered. Specifically, the linear regression models in Table 6.4 revealed significant correlations to how much of their relative attention was focused on the selected and non-selected options listed as "Avg answer fixation" and "Avg nonanswer fixation", respectively. This specific finding possibly indicates that the ETMP takes into account the validity of the Eye-Mind Theory [79]. Additionally, we also found a correlation between the ETMP and fixation on the appropriate answers in the game as indicated by the "Appropriate Fix" factor. This can be interpreted to mean that the ETMP also considers the accuracy of the mental schema that a player uses to determine his or her response though this interpretation is not entirely definitive. We also considered the possibility that rather this correlation may have occurred due to the low-score options being obviously socially unacceptable and easy to discard from consideration. This would then mean that the correct fixation was due in part to an understanding of general social norms rather than cultural competency. A targeted experiment with dialogue option sets that were more closely clustered in terms of appropriateness could resolve this interpretation. Finally, the ETMP was found to be strongly correlated to the player's explicit game score which we interpret as being that the players' explicit performances were likely attributable to cultural proficiency.

6.7 Limitations

The CCDTS dialogue structures were meant to facilitate a multitude of conversation tracks depending on the responses selected by the participants. Though this feature is ideal when considering realistic gameplay, it added complexity to the data structures and subsequently limited the sequential analysis that could be conducted since each player experienced different branches of the dialogue trees. For this reason, we were only able to effectively analyze the aggregated visual behavior metrics of each participant though we believe that a sequential analysis would provide insight into the specific metacognitive strategies employed by participants as well as learning rates.

6.8 Conclusion

In summary, this study introduces a novel eye-tracking measure of performance in social interaction serious games. We present a unique methodology by which players of DGBL systems can be assessed for not only their explicit in-game responses, but also for the correctness of the mental schema used to determine the selected response. This study also presents the results of an experiment whereby the ETMP was found to be correlated to gaze-behaviors indicative of specific cognitive processes and traditional measures of in-game performance. These findings contribute to the initial validation of the ETMP and motivate further research of the ETMP in other DGBL domains.

Chapter 7

Conclusions

"It is by no means an irrational fancy that, in a future existence, we shall look upon what we think our present existence, as a dream."

Edgar Allan Poe

This dissertation presents novel ways to develop and assess effective 3C DGBL systems using existing Cultural Dimension theory for content development, HMD VR as a channel for delivering 3C content, and leveraging gaze behavior as a means of assessing performance. In the following section, the author highlights and discusses the implications of this research both in the area of 3C DGBLs as well as DGBLs in general. Table 7.1 references the other studies in the literature with similar findings as those presented in this dissertation. The findings in this dissertation that have no other comparable studies are labelled as 'Novel'.

7.1 Engineering/Procedural Contributions

3C Game Review and Taxonomy

Through a thorough review of the literature, the author provides an organization and taxonomy of current research efforts. In doing so, the author highlights research and development trends in the area of 3C games and identifies knowledge gaps and research opportunities for future studies. The author of this dissertation explores several of the gaps that were identified in this taxonomy.

Cultural Simulation Design Process

This dissertation also presents a methodology by which existing Cultural Dimension Theory can be utilized to develop game content for 3C games. Several research efforts have been made to investigate various development guidelines and methodologies for creating effective 3C DGBLs, and many of them target high-level game artifacts such as agent models and ITS dynamics[98][118]. Although these artifacts are extremely relevant to the development of 3C games, this dissertation is focused on the specific conversational content that players must often navigate to achieve 3C game objectives. The methodology presented in this dissertation uses Hofstede's Cultural Dimensions to develop dimension-specific dialogue options to implicitly teach the existence of the cultural dimensions[67]. Additionally, this method provides a framework by which cultural content can be developed that specifically targets cultural dimensions that are the most disparate between the culture of the player and that of the 3C DGBL. In doing so, it is the hope that players can also gain a more general understanding of the cultural dimension being trained and its applicability across cultures. As a result, games designed using this methodology have the potential of having broader learning benefits beyond those of the specific culture for which a game is designed.

3C DGBLs Role-Playing Situational Judgment Test

The introduction of the role-playing SJT is one minor contribution of this dissertation to the assessment methodologies of 3C DGBL systems. To achieve a more realistic measure of 3C gains from DGBL systems, the author of this dissertation presents an evaluation methodology that involves real interactions between players and humans. These interactions are then evaluated for their cultural consistency and effectiveness by cultural experts. Although this was demonstrated in real-time face-to-face interactions, this methodology can also be executed virtually through video teleconferencing systems in cases in which cultural experts are not locally available.

7.2 Scientific/Technical Discovery Contributions

Alhough this research focuses on the effective development and assessment of 3C DGBLs, this dissertation presents findings that have implications for areas beyond the narrow scope of 3C DGBLs.

Regarding DGBLs designed for HMD VR, this research adds to the body of work about the efficacy of this highly-immersive and entertaining medium. The current body of research has addressed the efficacy of HMD VR versus traditional desktop channels in one off cases and the findings made are often generalized across all

TABLE 7.1: Comparison of Dissertation Findings and Other Studies

Dissertation Findings	**Similar Findings in Other Research (if any)**
Literature review and development of a taxonomy to identify gaps in the research of 3C Games	Novel
Development of a new method called the Cultural Dimension Design Process for generating Cultural Dimension Theory dialogue content for 3C DGBL systems	Novel
Demonstration of a game developed using the Cultural Dimension Design Process and the statistical motivation for its efficacy	Novel
Introduction and demonstration of the Role-Playing Situational Judgment Test to evaluate learning benefits in 3C DGBL systems	Novel
Development of an eye-tracking based measure of performance in DGBL systems	Just(1980[79]
Demonstration of the eye-tracking based measure of performance in an experiment showing the correlation of the novel measure against various other measures of performance	Glaholt(2011)[56], Shimojo(2003)[124], Duchowski(2002)[38]
Demonstration that HMD VR and traditional desktop variants of a DGBL result in insignificant performance differences	Freina(2016)[51], Freina(2015)[50], Tacket(2016)[130]
Discovery that gender is a significant factor when determining learning outcomes in immersive DGBL systems	Guadagno(2011)[57], Coffey(2017)[29], Bailenson(2008)[11], Yee(2006)[148]

domains. Although some of these claims may have merit, it is imperative for researchers in this area to begin to examine the efficacy of HMD VR regarding those domains that are advantaged by high levels of immersion.

In the case of 3C and other social skills, sensory acuity and situational awareness are often identified as attributes that contribute to high levels of effectiveness in social situations[25]. As such, it seems logical that greater levels of immersion would lead to positive results. The author of this research began to explore this question through 3C DGBLs, but it would seem prudent to examine similar questions in other social skills DGBLs to be able to make broader claims. Although the results in this dissertation reveal no performance differences between the two channels in general, they contributed to the growing trend of performance gains in HMD VR specifically for males specifically[29]. Although some theories have been presented to explain this interaction, none were found to have been explicitly experimented upon[10]. Given the large financial investments being made in the technical performance of HMD VR as well as the lack of specific guidance for the development of effective HMD VR-based DGBL systems, the findings of this dissertation suggest that further research is warranted to ensure that optimal DGBLs are developed.

Finally, the author of this dissertation presents and documents initial validation results for the use of physiological measurements as a means of assessing performance in DGBLs. Specifically, the author looks at a means of weighting fixation duration across dialogue option choices as a means of non-intrusively evaluating the decision-making process of players in a DGBL system. While eye-tracking metrics have been most often used as exploratory measures to understand various visual behaviors, the maturity of visual behavior theory has allowed for some eye-tracking measures to be used to assess performance. As such, given that most DGBLs continue to rely on explicit multiple-choice decisions, leveraging eye-tracking to interpret and evaluate decision-making in DGBL environments is promising. The results shown in this dissertation suggest that using eye-tracking as a means of evaluating performance may provide greater insight into individual skills than current methods of explicit assessment.

7.3 Future Works

Beyond on the limits of this dissertation, a number of extensions and investigations are proposed to be researched in future work.

1. The budding area of HMD VR continues to garner widespread support from companies like Facebook, Google, and Apple. The primary focus of these companies, however, is to increase the entertainment value of this new medium through advancements in resolution, video immersion, and audio immersion. As the DGBL community continues to adopt these technologies, it is imperative for future researchers to remain abreast with the learning value that can be achieved through these enhancements. Due to the high financial cost of HMD VR development, many of the cutting-edge capabilities of HMD VR games were not leveraged. These items include 360 audio, haptic interactions with game artifacts, gaze-controlled user interfaces, and gaze-responsive avatars. Many of these enhancements have the potential of greatly increasing the immersion of HMD VR systems greatly beyond the scope of this research. As these enhancements are brought to market, academic research efforts should be focused on evaluating their value related to learning benefits. Specifically, it

would be ideal for these research efforts to be targeted at individual enhancements to identify promising future developments as well as holistically to assess the comprehensive immersive effects of all the enhancements in-tandem. As the cost and availability of development tools becomes increasingly accessible, many of these enhancements will become increasingly financially feasible to evaluate.

2. In the area of eye-tracking in DGBLs, many research extensions and opportunities have become available. One major limitation of the research presented in this dissertation is that the time series effect of fixation duration on dialogue responses was not investigated due to limitations in the stimuli. Given that various eye-tracking theorists have presented research results that show evolutions in gaze behavior as knowledge and confidence change, it would seem prudent to investigate the same types of effects within DGBLs. Additionally, eye-tracking technology is being integrated into HMD VR systems by several companies. This presents a unique research opportunity to investigate whether the ETMP shows comparable or even more favorable results in immersive channels.

3. Regarding the use of Cultural Dimension theory in developing 3C DGBLs, the method presented in this research specifically leverages the findings of Geert Hofstede[67]. As modern cultural theorists begin to adapt their views from national cultures to blended cultures without national boundaries it may be necessary for future researchers to develop cultural content methodologies that do not rely on the definitions and limitation of ethnic or national cultures. As these theories continue to evolve and empirical evidence supporting these theories is presented in the literature, it is likely that future DGBL systems will need to rely on more adaptive and less culturally-restrictive methods of developing

cultural content for 3C DGBL systems.

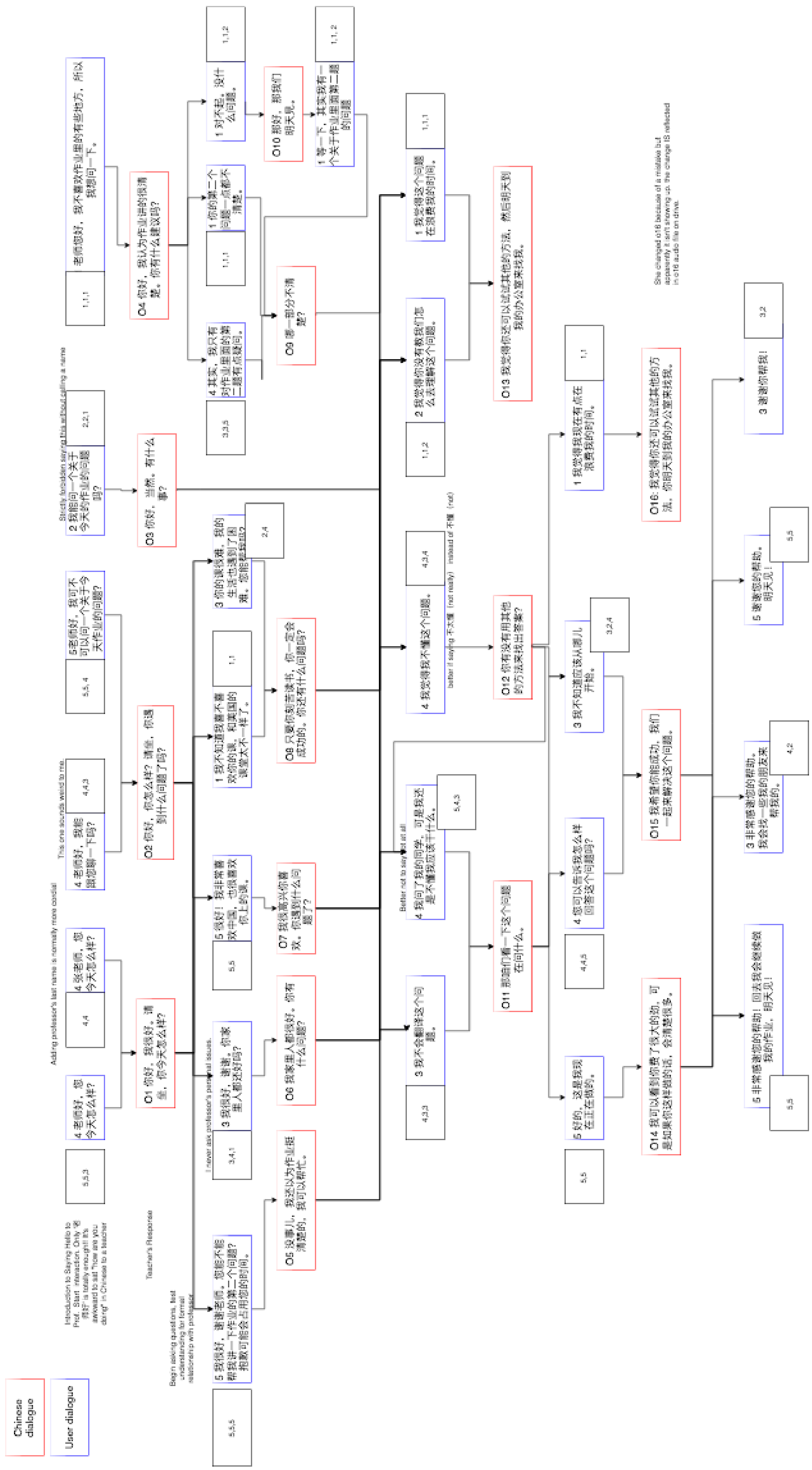
7.4 Finale

The results presented in this dissertation demonstrate how 3C DGBL systems can be more effectively developed and assessed by using existing Cultural theory, exercising modern immersive technologies, and leveraging physiological measures to gauge implicit actions. Given the massive growth in DGBLs in various domains, the research results suggest that the findings can potentially go beyond those of 3C-specific games. Although open questions continue to exist about means and methods of developing effective 3C DGBLs and DGBLs in general, the findings of this dissertation can be applied to current DGBL developments to give learners the most efficient and effective educational experiences possible.

Appendix A

Chinese Dialogue Tree for Office Hours Scenario

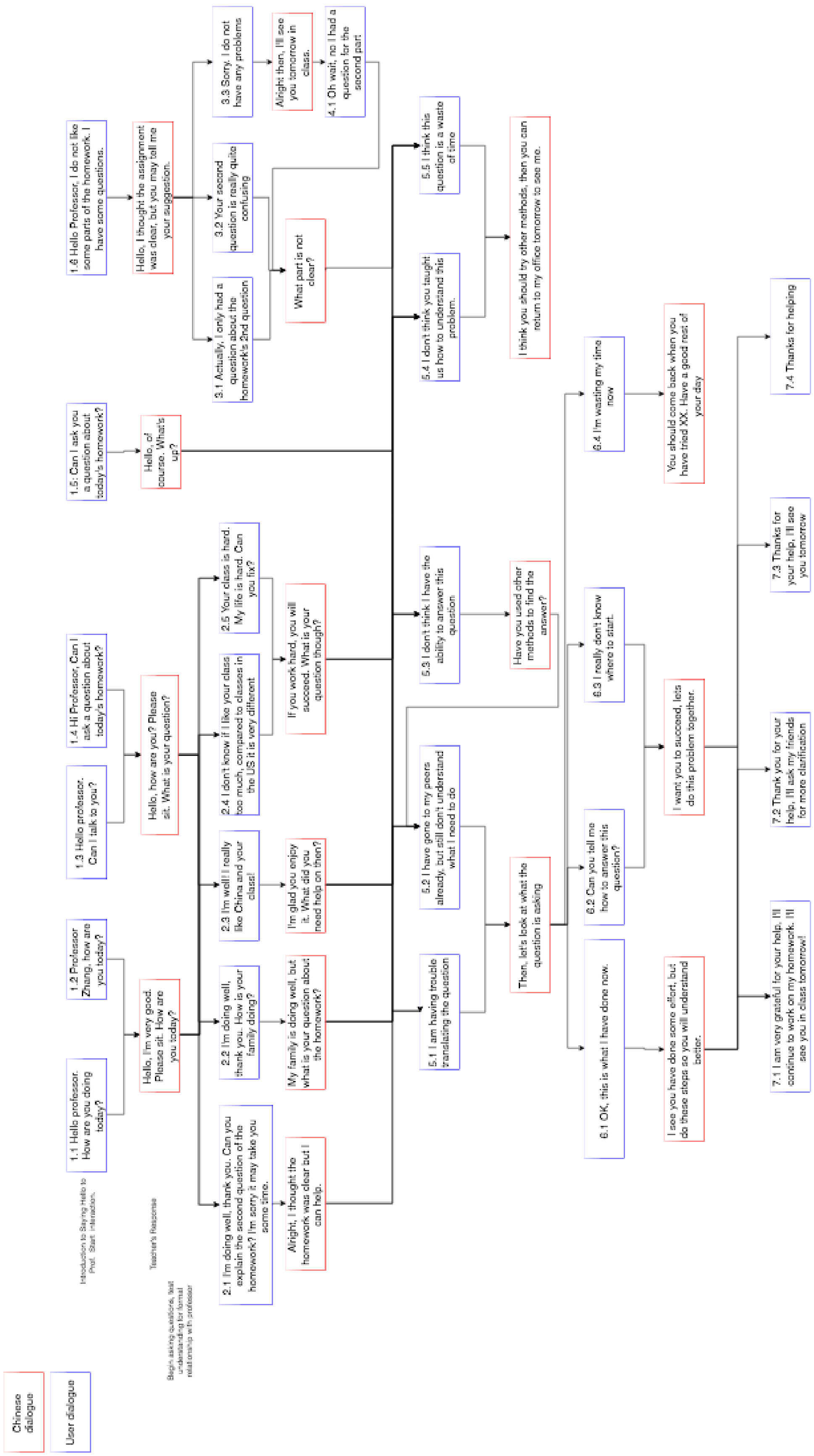
This appendix shows a full example of a Chinese dialogue tree for the Office Hours scenario.



Appendix B

English Dialogue Tree for Office Hours Scenario

This appendix shows a full example of an English dialogue tree for the Office Hours scenario.


Chinese dialogue
User dialogue
Introduction to Saying Hello to Prof. Start interaction.
Teacher's Response
Begin asking questions, test understanding for formal relationship with professor
1.1 Hello professor. How are you doing today?
1.2 Professor Zhang, how are you today?
1.3 Hello professor. Can I talk to you?
1.4 Hi Professor, Can I ask a question about today's homework?
1.5: Can I ask you a question about today's homework?
1.6 Hello Professor, I do not like some parts of the homework. I have some questions.
Hello, I'm very good. Please sit. How are you today?
Hello, how are you? Please sit. What is your question?
Hello, of course. What's up?
Hello, I thought the assignment was clear, but you may tell me your suggestion.
2.1 I'm doing well, thank you. Can you explain the second question of the homework? I'm sorry it may take you some time.
2.2 I'm doing well, thank you. How is your family doing?
2.3 I'm well! I really like China and your class!
2.4 I don't know if I like your class too much, compared to classes in the US it is very different
2.5 Your class is hard. My life is hard. Can you fix?
3.1 Actually, I only had a question about the homework's 2nd question
3.2 Your second question is really quite confusing
3.3 Sorry. I do not have any problems
Alright, I thought the homework was clear but I can help.
My family is doing well, but what is your question about the homework?
I'm glad you enjoy it. What did you need help on then?
If you work hard, you will succeed. What is your question though?
What part is not clear?
Alright then, I'll see you tomorrow in class.
4.1 Oh wait, no I had a question for the second part
5.1 I am having trouble translating the question
5.2 I have gone to my peers already, but still don't understand what I need to do
5.3 I don't think I have the ability to answer this question
5.4 I don't think you taught us how to understand this problem.
5.5 I think this question is a waste of time
Then, let's look at what the question is asking
Have you used other methods to find the answer?
I think you should try other methods, then you can return to my office tomorrow to see me.
6.1 OK, this is what I have done now.
6.2 Can you tell me how to answer this question?
6.3 I really don't know where to start.
6.4 I'm wasting my time now
I see you have done some effort, but do these steps so you will understand better.
I want you to succeed, lets do this problem together.
You should come back when you have tried XX. Have a good rest of your day
7.1 I am very grateful for your help. I'll continue to work on my homework. I'll see you in class tomorrow!
7.2 Thank you for your help, I'll ask my friends for more clarification
7.3 Thanks for your help, I'll see you tomorrow
7.4 Thanks for helping

Appendix C

Actor Post-Test Dossier

Mindset: The choice of whether or not the objective is reached is based on how culturally appropriate the participant behaves. You will have some initial interactions laid out as an example for you to start you off, but then the conversation should differ based on the participant's conversation decisions. As the participant makes dialogue decisions, you should alter your responses depending on whether the participant is being culturally appropriate or not. Actor Post-Test Participant Scenario: Asking a Chinese boss for a raise

Time Limit: You have four minutes to review this dossier. The conversation will be about ten minutes

Participant's Objective: To get a raise without upsetting the boss

Scenario Description (for the Participant): "You have been working as a roofing contractor for the same Chinese company for three years without a pay raise. After much careful consideration, you have decided to ask your boss for a raise. Because you are a contractor, the boss will see you as subordinate. In order to receive the raise, you must demonstrate that you are worthy of a raise while still remaining sensitive to their higher position and attempting to demonstrate respect. "

Actor's Role: Your role is to be the Chinese boss that the participant is trying to get a raise from. You are a strict boss but want to reward those who do exceptional work. This, though, means that not many people in your company get raises. You must be convinced that the participant deserves a raise without the participant acting inappropriately. Chinese hierarchies can be more rigid by not allowing much decision-making authority to lower leaders. As the boss may not have the authority to grant the participant a raise, respond to the participant's request for a raise by introducing the prospect of having to first discuss things with your supervisor to see how they react (i.e. are they accommodating for this cultural reality?) Because the Chinese tend to have a long-term orientation, you are concerned about the future of your company and want to increase your advertising sources. In order to see how the participant responds to this common phenomenon among the Chinese, ask the participant if he/she would be willing to place signs for your company in the lawns of their customers on the day of the job. Explain that this will help the company's long-term success. Ideally, the participant will recognize your tendency toward long-term orientation and will be accomodating.

Actor Personality Traits: Strict, Fair, Type A personality

Actor Background: You are a Chinese boss that has been working for this company for a decade and a half. You know the company very well. You look for hard workers in the company but are stringent on who you hire. You have only given raises to people who have gone above and beyond in their positions.

Beginning Dialogue Example (Actor): <You begin the conversation> Greeting: "Hello, come in. Why did you want to meet with me today? Is there a problem relating to your work?"

<Participant responds>

Response 1.1: No, sir/ma'am. I think that we should discuss my current status in this company. I believe that I should receive an increase in salary. Response 1.2: No, sir/ma'am. I want to talk about how I have been overlooked in this company. I believe that I have been mistreated and deserve a raise. Response 1.3: No, sir/ma'am. I actually wanted to discuss something with you. I was hoping to have an open conversation with you about potentially increasing my salary.

<You select a response> (If Response 1.1 is chosen): Do not come into my office and expect to get something out from me. Remember that I am superior. Why should you receive a raise? (If Response 1.2 is chosen): What do you mean? You need to justify to me why you deserve a raise (If Response 1.3 is chosen): Sure, let us discuss this. Why do you think you deserve a raise?

<Participant responds> Response 2.1: I have been working here a long time and done a lot. I believe that I am due a raise because of this. Response 2.2: I have done a lot of good things for the company. I have lead many projects, brought in a lot of business for the company, and have taken a lot of extra work that was not assigned to me. Response 2.3: I'm the best worker you have in this company. I have done a lot for you and this company but I believe that I am always overlooked. I deserve to be paid for the value I bring.

<You select a response> (If Response 2.1 was selected): I can't give you a raise just because you think you deserve one. I need to discuss with higher management. (If Response 2.2 was selected): I have seen the work ethic you bring to the table. There is potential for you to receive a raise, I believe. I want to see what upper management says as well. (If Response 2.3 was selected): You should be humble about your accomplishments. A lot of people in this company do good work. Furthermore, upper management has a say, as well, on who receives a raise.

<Participant Responds> Response 3.1: I can show them all the work I have been doing over the past couple months. I believe they would be very impressed. Response 3.2: Aren't you the one, though, who has a final say on who gets a raise? Why do we need to discuss this with upper management? Response 3.3: You can ask anyone in the office, including upper management, about my work ethic. I work harder than almost anyone.

<You respond - you will now go off script>

Appendix D

Participant Post-Test Dossier

Participant Post-Test

Scenario: Asking a Chinese boss for a raise

Objective: To get a raise without upsetting your boss

Time Limit: You have four minutes to review this dossier. The conversation will be about ten minutes

Scenario Description: You have been working as a roofing contractor for the same Chinese company for three years without a pay raise. After much careful consideration, you have decided to ask your boss for a raise. Because you are a contractor, the boss will see you as subordinate. In order to receive the raise, you must demonstrate that you are worthy of a raise while still remaining sensitive to their higher position and attempting to demonstrate respect.

Participant Background: Three years ago, you moved to China to work in the Chinese office of your company. You thought that this would be a great way to move up in the company but your position has been stagnant for a while. You have won numerous Employee of the Month awards, earned the highest productivity rate in the office last year, and have filled in for higher management numerous times. You believe that your current salary does not reflect the amount of work that you have been doing and believe you should receive a raise.

Actor's Role: The actor's role is to act as the Chinese boss. You know the boss to be strict but fair to their employees.

Beginning Dialogue Example (Participant):

<Actor starts conversation>

<You select a response>

Response 1.1: No, sir/ma'am. I think that we should discuss my current status in this company. I believe that I should receive an increase in salary. Response 1.2: No, sir/ma'am. I want to talk about how I have been overlooked in this company. I believe that I have been mistreated and deserve a raise. Response 1.3: No, sir/ma'am. I actually wanted to discuss something with you. I was hoping to have an open conversation with you about potentially increasing my salary.

<Actor responds>

<You select a response> Response 2.1: I have been working here a long time and done a lot. I believe that I am due a raise because of this. Response 2.2: I have done a lot of good things for the company. I have lead many projects, brought in a lot of business for the company, and have taken a lot of extra work that was not assigned to me. Response 2.3: I'm the best worker you have in this company. I have done a lot for you and this company but I believe that I am always overlooked. I deserve to be paid for the value I bring.

<You select a response> Response 3.1: I can show them all the work I have been doing over the past couple months. I believe they would be very impressed. Response 3.2: Aren't you the one, though, who has a final say on who gets a raise? Why do we

need to discuss this with upper management? Response 3.3: You can ask anyone in the office, including upper management, about my work ethic. I work harder than almost anyone.

<Actor responds - goes off script> <You respond - go off script>

Appendix E

Example Post-Test Rubric

Item	Rating of 2	Rating of 1	Rating of 0
Presence of Introduction	Greets and opens with 'relationship building'	Says hello but immediately proceeds to the issue	Does not greet and opens with the problem
Content of Introduction	Introduction is culturally sensitive/ appropriate	Introduction is culturally acceptable	Introduction is culturally awkward/ inappropriate
Acknowledgement of Hierarchy	Is respectful to the boss in light of boss' seniority	Moderately respectful to the boss	Does not engage respectfully with the boss
Explains Desire for Raise	Clearly explains the issue	Hints at the issue - but no direct mention	Does not address wanting a raise
Flexible with Company Hierarchical Structure	Understands the need for a higher boss' approval of raise	Accepts that boss must receive permission to grant a raise	Not accepting of the fact that the boss must receive permission to grant a raise
Asks for Opinion/Thoughts	Inquires why they haven't gotten a raise and considers input in response	Inquires why they haven't gotten a raise	Does not ask about reason for not getting a raise
Explains Consequences	Explains impact of not getting a raise in so long	Hints at impact/potential consequences of not getting a raise in so long	Does not acknowledge potential consequences
Flexible with Marketing Objectives of Boss	Understanding of the boss' request for marketing assistance	Tolerant of the boss' request for marketing assistance	Does not agree to the request for marketing assistance being placed
Conclusion/ Farewell	Farewell is culturally sensitive/ appropriate	Farewell is culturally acceptable	Farewell is culturally awkward/ inappropriate

Bibliography

[1] S. Sam Adkins. *The 2017-2022 Global Game-based Learning Market*. Tech. rep. Serious Play Conference 2017, 2017.

[2] Hirotugu Akaike. “A new look at the statistical model identification”. In: *IEEE transactions on automatic control* 19.6 (1974), pp. 716–723.

[3] Hirotugu Akaike. “Factor analysis and AIC”. In: *Psychometrika* 52.3 (1987), pp. 317–332.

[4] Wadee S Alhalabi. “Virtual reality systems enhance students’ achievements in engineering education”. In: *Behaviour & Information Technology* 35.11 (2016), pp. 919–925.

[5] Ilan Alon et al. “Business Cultural Intelligence Quotient: A Five-Country Study”. In: *Thunderbird International Business Review* 60.3 (2018), pp. 237–250.

[6] Brian An et al. “Cultural Dimension Theory Based Simulations for US Army Personnel”. In: *International Conference on Social Computing, Behavioral-Cultural Modeling and Prediction and Behavior Representation in Modeling and Simulation*. Springer. 2017, pp. 65–70.

[7] Philip H Anderson and Leigh Lawton. “Student motivation to study abroad and their intercultural development”. In: *Frontiers: The Interdisciplinary Journal of Study Abroad* (2015).

[8] Soon Ang, Linn Van Dyne, and Christine Koh. “Personality correlates of the four-factor model of cultural intelligence”. In: *Group & Organization Management* 31.1 (2006), pp. 100–123.

[9] Leonard A Annetta. "The "I's" have it: A framework for serious educational game design." In: *Review of General Psychology* 14.2 (2010), p. 105.

[10] Lynna J Ausburn et al. "A cross-case analysis of gender issues in desktop virtual reality learning environments". In: (2009).

[11] Jeremy N Bailenson et al. "The use of immersive virtual reality in the learning sciences: Digital transformations of teachers, students, and social context". In: *The Journal of the Learning Sciences* 17.1 (2008), pp. 102–141.

[12] Azadeh Bashiri, Marjan Ghazisaeedi, and Leila Shahmoradi. "The opportunities of virtual reality in the rehabilitation of children with attention deficit hyperactivity disorder: a literature review". In: *Korean journal of pediatrics* 60.11 (2017), pp. 337–343.

[13] Theo Bastiaens, Lincoln Wood, and Torsten Reiners. "New landscapes and new eyes: The role of virtual world design for supply chain education". In: *Ubiquitous Learning* 6.1 (2014), pp. 37–49.

[14] Christopher Bayer et al. "The SIMPLE methodology for developing and evaluating cross-cultural virtual training systems SIEDS 2017". In: *Systems and Information Engineering Design Symposium (SIEDS), 2017*. IEEE. 2017, pp. 289–293.

[15] Francesco Bellotti et al. "Assessment in and of serious games: an overview". In: *Advances in Human-Computer Interaction* 2013 (2013), p. 1.

[16] Milton J Bennett. "A developmental approach to training for intercultural sensitivity". In: *International journal of intercultural relations* 10.2 (1986), pp. 179–196.

[17] Milton J Bennett. "Becoming interculturally competent". In: *Toward multiculturalism: A reader in multicultural education* 2 (2004), pp. 62–77.

[18] Dharm Bhawuk and Richard Brislin. "Cross-cultural training: a review". In: *Applied Psychology* 49.1 (2000), pp. 162–191.

[19] Melanie Bloom and Arturo Miranda. "Intercultural sensitivity through short-term study abroad". In: *Language and Intercultural Communication* 15.4 (2015), pp. 567–580.

[20] Ryan Bradley and Nigel Newbutt. "Autism and Virtual Reality Head-Mounted Displays: A State of the Art Systematic Review". In: *Journal of Enabling Technologies* 11 (2018).

[21] Joost JLE Bücker and Hubert Korzilius. "Developing cultural intelligence: assessing the effect of the Ecotonos cultural simulation game for international business students". In: *The International Journal of Human Resource Management* 26.15 (2015), pp. 1995–2014.

[22] Andreas Bulling et al. "Eye movement analysis for activity recognition using electrooculography". In: *IEEE transactions on pattern analysis and machine intelligence* 33.4 (2011), pp. 741–753.

[23] Luis Cerezo, Allison Caras, and Ronald P Leow. "The effectiveness of guided induction versus deductive instruction on the development of complex Spanish gustar structures: An analysis of learning outcomes and processes". In: *Studies in Second Language Acquisition* 38.2 (2016), pp. 265–291.

[24] Guo-Ming Chen and William J Starosta. "The development and validation of the intercultural sensitivity scale". In: (2000).

[25] Chi-Yue Chiu et al. "Cross-cultural competence: Theory, research, and application". In: (2013).

[26] Chi-Yue Chiu et al. *Cross-cultural competence: Theory, research, and application*. 2013.

[27] Roy YJ Chua and Kok Yee Ng. "Not Just How Much You Know: Interactional Effect of Cultural Knowledge and Metacognition on Creativity in a Global Context". In: *Management and Organization Review* 13.2 (2017), pp. 281–300.

[28] Amy Jo Coffey et al. "New media environments' comparative effects upon intercultural sensitivity: A five-dimensional analysis". In: *International Journal of Intercultural Relations* 37.5 (2013), pp. 605–627.

[29] Amy Jo Coffey et al. "The efficacy of an immersive 3D virtual versus 2D web environment in intercultural sensitivity acquisition". In: *Educational Technology Research and Development* 65.2 (2017), pp. 455–479.

[30] Thomas M Connolly et al. "A systematic literature review of empirical evidence on computer games and serious games". In: *Computers & Education* 59.2 (2012), pp. 661–686.

[31] Thomas D Cook, Donald Thomas Campbell, and William Shadish. *Experimental and quasi-experimental designs for generalized causal inference*. Houghton Mifflin Boston, 2002.

[32] François Courtemanche et al. "Activity recognition using eye-gaze movements and traditional interactions". In: *Interacting with Computers* 23.3 (2011), pp. 202–213.

[33] Chris Crawford. *Chris Crawford on interactive storytelling*. New Riders, 2012.

[34] Nicole L Cundiff, Joel T Nadler, and Alicia Swan. "The influence of cultural empathy and gender on perceptions of diversity programs". In: *Journal of Leadership & Organizational Studies* 16.1 (2009), pp. 97–110.

[35] Joao Dias, Samuel Mascarenhas, and Ana Paiva. "Fatima modular: Towards an agent architecture with a generic appraisal framework". In: *Emotion modeling*. Springer, 2014, pp. 44–56.

[36] Lisa Dieker et al. "Virtual classrooms: STAR simulator". In: *New Learning Technology SALT* 4 (2007), pp. 1–22.

[37] Sidney D'Mello et al. "Gaze tutor: A gaze-reactive intelligent tutoring system". In: *International Journal of human-computer studies* 70.5 (2012), pp. 377–398.

[38] Andrew T Duchowski. “A breadth-first survey of eye-tracking applications”. In: *Behavior Research Methods, Instruments, & Computers* 34.4 (2002), pp. 455–470.

[39] Nick C Ellis. “Implicit and explicit learning of languages”. In: (1994).

[40] Lilli Engle and John Engle. “Assessing language acquisition and intercultural sensitivity development in relation to study abroad program design.” In: *Frontiers: The interdisciplinary journal of study abroad* 10 (2004), pp. 219–236.

[41] Linda Eriksson. *The Effectiveness of Modified Inductive Versus Deductive Teaching: A case study on word order amongst a group of English as a foreign language learners*. 2014.

[42] Tony Fang. “Yin Yang: A new perspective on culture”. In: *Management and organization Review* 8.1 (2012), pp. 25–50.

[43] Denise Rotondo Fernandez et al. “Hofstede’s country classification 25 years later”. In: *The Journal of social psychology* 137.1 (1997), pp. 43–54.

[44] Paul A Fishwick et al. “An experimental design and preliminary results for a cultural training system simulation”. In: *Simulation Conference (WSC), Proceedings of the 2010 Winter*. IEEE. 2010, pp. 799–810.

[45] Paul A Fishwick et al. “Simulating culture: an experiment using a multi-user virtual environment”. In: *Simulation Conference, 2008. WSC 2008. Winter*. IEEE. 2008, pp. 786–794.

[46] Joseph L Fleiss and Jacob Cohen. “The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability”. In: *Educational and psychological measurement* 33.3 (1973), pp. 613–619.

[47] *Foreign Area Officer Program*. `https://www.dliflc.edu/home/about/fao/`. Accessed: 2018-05-16.

[48] Sandra M Fowler and Margaret D Pusch. "Intercultural simulation games: A review (of the United States and beyond)". In: *Simulation & Gaming* 41.1 (2010), pp. 94–115.

[49] Daniel Freeman et al. "Virtual reality in the assessment, understanding, and treatment of mental health disorders". In: *Psychological medicine* 47.14 (2017), pp. 2393–2400.

[50] Laura Freina and Michela Ott. "A literature review on immersive virtual reality in education: state of the art and perspectives". In: *The International Scientific Conference eLearning and Software for Education*. Vol. 1. " Carol I" National Defence University. 2015, p. 133.

[51] Laura Freina et al. "Immersion's impact on performance in a spatial reasoning task". In: *International Conference on Games and Learning Alliance*. Springer. 2016, pp. 211–220.

[52] Josef Froschauer et al. "Design and evaluation of a serious game for immersive cultural training". In: *Virtual Systems and Multimedia (VSMM), 2010 16th International Conference on*. IEEE. 2010, pp. 253–260.

[53] Josef Froschauer et al. "Designing socio-cultural Learning Games: Challenges and lessons learned". In: *Information Society (i-Society), 2011 International Conference on*. IEEE. 2011, pp. 56–61.

[54] Maite Frutos-Pascual and Begonya Garcia-Zapirain. "Assessing visual attention using eye tracking sensors in intelligent cognitive therapies based on serious games". In: *Sensors* 15.5 (2015), pp. 11092–11117.

[55] Jessica A Gallus et al. *Cross-cultural competence in the department of defense: an annotated bibliography*. Tech. rep. DTIC Document, 2014.

[56] Mackenzie G Glaholt and Eyal M Reingold. "Eye movement monitoring as a process tracing methodology in decision making research." In: *Journal of Neuroscience, Psychology, and Economics* 4.2 (2011), p. 125.

[57] Rosanna E Guadagno et al. "Even in virtual environments women shop and men build: A social role perspective on Second Life". In: *Computers in Human Behavior* 27.1 (2011), pp. 304–308.

[58] Edward Twitchell Hall. "The hidden dimension". In: (1966).

[59] Edward Twitchell Hall et al. *The silent language*. Vol. 3. Doubleday New York, 1959.

[60] Mitchell R Hammer, William B Gudykunst, and Richard L Wiseman. "Dimensions of intercultural effectiveness: An exploratory study". In: *International Journal of Intercultural Relations* 2.4 (1979), pp. 382–393.

[61] Roger Harrison and Richard L Hopkins. "The design of cross-cultural training: An alternative to the university model". In: *The Journal of applied behavioral science* 3.4 (1967), pp. 431–460.

[62] Mary Hegarty, Richard E Mayer, and Carolyn E Green. "Comprehension of arithmetic word problems: Evidence from students' eye fixations." In: *Journal of Educational Psychology* 84.1 (1992), p. 76.

[63] Steven J Heine. *Cultural Psychology: Third International Student Edition*. WW Norton & Company, 2015.

[64] Steven J Heine et al. "What's wrong with cross-cultural comparisons of subjective Likert scales?: The reference-group effect." In: *Journal of personality and social psychology* 82.6 (2002), p. 903.

[65] Randall W Hill Jr et al. *Pedagogically structured game-based training: Development of the ELECT BiLAT simulation*. Tech. rep. UNIVERSITY OF SOUTHERN CALIFORNIA MARINA DEL REY CA INST FOR CREATIVE TECHNOLOGIES, 2006.

[66] Geert Hofstede. "Asian management in the 21st century". In: *Asia pacific journal of management* 24.4 (2007), pp. 411–420.

[67] Geert Hofstede. "Dimensionalizing cultures: The Hofstede model in context". In: *Online readings in psychology and culture* 2.1 (2011), p. 8.

[68] Geert H Hofstede and Geert Hofstede. *Culture's consequences: Comparing values, behaviors, institutions and organizations across nations*. Sage, 2001.

[69] Kristiina Holm, Petri Nokelainen, and Kirsi Tirri. "Relationship of gender and academic achievement to Finnish students' intercultural sensitivity". In: *High Ability Studies* 20.2 (2009), pp. 187–200.

[70] Isabelle Hupont et al. "How do new visual immersive systems influence gaming QoE? A use case of serious gaming with Oculus Rift". In: *Quality of Multimedia Experience (QoMEX), 2015 Seventh International Workshop on*. IEEE. 2015, pp. 1–6.

[71] Dušan Jan et al. "A computational model of culture-specific conversational behavior". In: *International Workshop on Intelligent Virtual Agents*. Springer. 2007, pp. 45–56.

[72] Natasha Jaques et al. "Predicting affect from gaze data during interaction with an intelligent tutoring system". In: *International Conference on Intelligent Tutoring Systems*. Springer. 2014, pp. 29–38.

[73] Brent K Jesiek, Yi Shen, and Yating Haller. "Cross-Cultural Competence: A Comparative Assessment of EngineeringStudents". In: *International Journal of Engineering Education* 28.1 (2012), p. 144.

[74] W Lewis Johnson. "Serious use of a serious game for language learning". In: *Frontiers in Artificial Intelligence and Applications* 158 (2007), p. 67.

[75] W Lewis Johnson and Andre Valente. "Tactical language and culture training systems: Using AI to teach foreign languages and cultures". In: *AI Magazine* 30.2 (2009), p. 72.

[76] W Lewis Johnson et al. “The Virtual Cultural Awareness Trainer (VCAT): Joint Knowledge Online’s (JKO’s) solution to the individual operational culture and language training gap”. In: *Proceedings of ITEC*. Clarion Events London, UK. 2011.

[77] Michael L Jones. “Hofstede-culturally questionable?” In: (2007).

[78] E Józsa and BP Hámornik. “Find the difference!: eye tracking study on information seeking behavior using an online game”. In: (2011).

[79] Marcel A Just and Patricia A Carpenter. “A theory of reading: From eye fixations to comprehension.” In: *Psychological review* 87.4 (1980), p. 329.

[80] Akira Kageyama and Asako Tomiyama. “Visualization framework for CAVE virtual reality systems”. In: *International Journal of Modeling, Simulation, and Scientific Computing* 7.04 (2016), p. 1643001.

[81] Irene Katsouri et al. “Visualizing and assessing hypotheses for marine archaeology in a VR CAVE environment”. In: *Journal on Computing and Cultural Heritage (JOCCH)* 8.2 (2015), p. 10.

[82] Theodore D Kemper. *Status, power and ritual interaction: A relational reading of Durkheim, Goffman and Collins*. Routledge, 2016.

[83] Ebrahim Khodadady and Shima Ghahari. “Validation of the Persian cultural intelligence scale and exploring its relationship with gender, education, travelling abroad and place of living”. In: *Global Journal of Human Social Science* 11 (2011).

[84] Michael D Kickmeier-Rust, Eva Hillemann, and Dietrich Albert. “Tracking the UFO’s paths: using eye-tracking for the evaluation of serious games”. In: *International Conference on Virtual and Mixed Reality*. Springer. 2011, pp. 315–324.

[85] Kristian Kiili, Harri Ketamo, and Michael D Kickmeier-Rust. "Evaluating the usefulness of Eye Tracking in Game-based Learning". In: *International Journal of Serious Games* 1.2 (2014).

[86] Roger E Kirk. *Experimental design*. Wiley Online Library, 1982.

[87] Felix Kistler et al. "Traveller: An interactive cultural training system controlled by user-defined body gestures". In: *IFIP Conference on Human-Computer Interaction*. Springer. 2013, pp. 697–704.

[88] Florence R Kluckhohn and Fred L Strodtbeck. "Variations in value orientations." In: (1961).

[89] Richard Lambert. "Language and intercultural competence". In: *Striving for the Third Place: Intercultural competence through language education* (1999), pp. 65–72.

[90] H Chad Lane et al. "Learning intercultural communication skills with virtual humans: Feedback and fidelity." In: *Journal of Educational Psychology* 105.4 (2013), p. 1026.

[91] Kwok Leung et al. "Culture and international business: Recent advances and their implications for future research". In: *Journal of International Business Studies* 36.4 (2005), pp. 357–378.

[92] Gerald L Lohse and Eric J Johnson. "A comparison of two process tracing methods for choice tasks". In: *Organizational Behavior and Human Decision Processes* 68.1 (1996), pp. 28–43.

[93] Guido Makransky, Thomas S Terkildsen, and Richard E Mayer. "Adding immersive virtual reality to a science lab simulation causes more presence but less learning". In: *Learning and Instruction* (2017).

[94] Ammar Maleki and Martin de Jong. "A proposal for clustering the dimensions of national culture". In: *Cross-Cultural Research* 48.2 (2014), pp. 107–143.

[95] Bertrand Marne et al. "The six facets of serious game design: a methodology enhanced by our design pattern library". In: *European Conference on Technology Enhanced Learning*. Springer. 2012, pp. 208–221.

[96] Samuel Mascarenhas et al. "Can I ask you a favour?: a relational model of socio-cultural behaviour". In: *Proceedings of the 2013 international conference on Autonomous agents and multi-agent systems*. International Foundation for Autonomous Agents and Multiagent Systems. 2013, pp. 1335–1336.

[97] Samuel Mascarenhas et al. "How should I say this? Agents with culturally-appropriate verbal communication styles". In: (2011).

[98] Samuel Mascarenhas et al. "Modeling culture in intelligent virtual agents". In: *Autonomous Agents and Multi-Agent Systems* 30.5 (2016), pp. 931–962.

[99] David Matsumoto and Hyisung C Hwang. "Assessing cross-cultural competence: A review of available tests". In: *Journal of cross-cultural psychology* 44.6 (2013), pp. 849–873.

[100] Chris McCollum et al. "Developing an immersive, cultural training system". In: *The Interservice/Industry Training, Simulation and Education Conference(I/ITSEC)*. 2004, p. 2004.

[101] Gary McGuire, Mrs Gail McGinn, and Mrs Nancy Weaver. *Developing and Managing Cross-Cultural Competence within the Department of Defense: Recommendations and Learning and Assessment*. `https://www.deomi.org/culturalreadines` `documents/racca_wg_sg2_workshop_report.pdf`. 2008.

[102] Merriam-Webster Online. *Merriam-Webster Online Dictionary*. 2017. URL: `http://www.merriam-webster.com`.

[103] José M Mestre et al. "Emotional intelligence and social and academic adaptation to school". In: *Psicothema* 18 (2006).

[104] David R Michael and Sandra L Chen. *Serious games: Games that educate, train, and inform*. Muska & Lipman/Premier-Trade, 2005.

[105] Snejina Michailova and Kate Hutchings. “National cultural influences on knowledge sharing: A comparison of China and Russia”. In: *Journal of Management Studies* 43.3 (2006), pp. 383–405.

[106] Konstantin Mitgutsch and Narda Alvarado. “Purposeful by design?: a serious game design assessment framework”. In: *Proceedings of the International Conference on the foundations of digital games*. ACM. 2012, pp. 121–128.

[107] Anne Moenning et al. “Developing avatars to improve cultural competence in US soldiers”. In: *Systems and Information Engineering Design Symposium (SIEDS), 2016 IEEE*. IEEE. 2016, pp. 148–152.

[108] Shira Mor, Michael W Morris, and Johann Joh. “Identifying and training adaptive cross-cultural management skills: The crucial role of cultural metacognition”. In: *Academy of Management Learning & Education* 12.3 (2013), pp. 453–475.

[109] Michela Mortara et al. “Evaluating the effectiveness of serious games for cultural awareness: the Icura user study”. In: *International Conference on Games and Learning Alliance*. Springer. 2013, pp. 276–289.

[110] Kalervo Oberg. “Cultural shock: Adjustment to new cultural environments”. In: *Practical anthropology* 7.4 (1960), pp. 177–182.

[111] Ernest H O’Boyle et al. “The relation between emotional intelligence and job performance: A meta-analysis”. In: *Journal of Organizational Behavior* 32.5 (2011), pp. 788–818.

[112] R Michael Paige, Andrew D Cohen, and Rachel L Shively. “Assessing the Impact of a Strategies-Based Curriculum on Language and Culture Learning Abroad.” In: *Frontiers: The interdisciplinary journal of study abroad* 10 (2004), pp. 253–276.

[113] Yaotian Pan, Julie A Rowney, and Mark F Peterson. “The structure of Chinese cultural traditions: An empirical study of business employees in China”. In: *Management and Organization Review* 8.1 (2012), pp. 77–95.

[114] Ekaterina Prasolova-Førland et al. “Training cultural awareness in military operations in a virtual Afghan village: A methodology for scenario development”. In: *System Sciences (HICSS), 2013 46th Hawaii International Conference on*. IEEE. 2013, pp. 903–912.

[115] Louise J Rasmussen, Winston R Sieck, and Jasmine L Duran. *A model of culture-general competence for education and training: Validation across services and key specialties*. Tech. rep. Global Cognition Yellow Springs, 2016.

[116] Elaine M Raybourn. “Applying simulation experience design methods to creating serious game-based adaptive training systems”. In: *Interacting with computers* 19.2 (2007), pp. 206–214.

[117] Elaine M Raybourn. “Applying simulation experience design methods to creating serious game-based adaptive training systems”. In: *Interacting with computers* 19.2 (2007), pp. 206–214.

[118] Elaine M Raybourn. “Intercultural competence game that fosters metacognitive agility and reflection”. In: *International Conference on Online Communities and Social Computing*. Springer. 2009, pp. 603–612.

[119] Arthur S Reber. “Implicit learning and tacit knowledge.” In: *Journal of experimental psychology: General* 118.3 (1989), p. 219.

[120] Tony Renshaw, Richard Stevens, and Paul D Denton. “Towards understanding engagement in games: An eye-tracking study”. In: *On the Horizon* 17.4 (2009), pp. 408–420.

[121] Sérgio Hortas Rodrigues et al. “A process model of empathy for virtual agents”. In: *Interacting with Computers* 27.4 (2014), pp. 371–391.

[122] Hafiz Muhammad Abdullah Shaiq et al. “Why not everybody loves Hofstede? What are the alternative approaches to study of culture”. In: *European Journal of Business and Management* 3.6 (2011), pp. 101–111.

[123] Thomas B Sheridan. “Musings on telepresence and virtual presence”. In: *Presence: Teleoperators & Virtual Environments* 1.1 (1992), pp. 120–126.

[124] Shinsuke Shimojo et al. “Gaze bias both reflects and influences preference”. In: *Nature neuroscience* 6.12 (2003), pp. 1317–1322.

[125] Julius Sim and Chris C Wright. “The kappa statistic in reliability studies: use, interpretation, and sample size requirements”. In: *Physical therapy* 85.3 (2005), pp. 257–268.

[126] Jeffrey Smith. “U.S. Military Admits Major Mistakes in Iraq and Afghanistan”. In: *The Atlantic* (2012). URL: https://www.theatlantic.com/international/archive/2012/06/us-military-admits-major-mistakes-in-iraq-and-afghanistan/258339/.

[127] Steven Solomon et al. *A language for modeling cultural norms, biases and stereotypes for human behavior models*. Tech. rep. UNIVERSITY OF SOUTHERN CALIFORNIA MARINA DEL REY CA INST FOR CREATIVE TECHNOLOGIES, 2008.

[128] Kaveri Subrahmanyam and Patricia M Greenfield. “Effect of video game practice on spatial skills in girls and boys”. In: *Journal of applied developmental psychology* 15.1 (1994), pp. 13–32.

[129] Eric A Surface, Erich C Dierdorff, and Aaron M Watson. “Special operations language training software measurement of effectiveness study: Tactical Iraqi study final report”. In: *Special Operations Forces Language Office, Tampa, FL, USA* (2007).

[130] Jeffrey Tackett. “Using a 3D immersive environment to study signal flow in music technology”. PhD thesis. Colorado Technical University, 2016.

[131] Marjon Tammenga-Helmantel et al. "Comparing inductive and deductive grammatical instruction in teaching German as a foreign language in Dutch classrooms". In: *System* 63 (2016), pp. 101–114.

[132] Sibel Ayper Tasdemir and Ekaterina Prasolova-Førland. "Visualizing Afghan Culture in a Virtual Village for Training Cultural Awareness in Military Settings". In: *Information Visualisation (IV), 2014 18th International Conference on*. IEEE. 2014, pp. 256–261.

[133] Harry C Triandis. "Culture specific assimilators". In: *Intercultural sourcebook: Cross-cultural training methods* 1 (1995), pp. 179–186.

[134] Fons Trompenaars and Charles Hampden-Turner. *Riding the waves of culture: Understanding diversity in global business*. Nicholas Brealey Publishing, 2011.

[135] Meng-Jung Tsai et al. "Visual attention for solving multiple-choice science problem: An eye-tracking analysis". In: *Computers & Education* 58.1 (2012), pp. 375–385.

[136] Linn Van Dyne, Soon Ang, and Christine Koh. "Development and validation of the CQS". In: *Handbook of cultural intelligence* (2008), pp. 16–40.

[137] Linn Van Dyne et al. "Sub-dimensions of the four factor model of cultural intelligence: Expanding the conceptualization and measurement of cultural intelligence". In: *Social and personality psychology compass* 6.4 (2012), pp. 295–313.

[138] Séverine Vogel et al. "Effectiveness of a guided inductive versus a deductive approach on the learning of grammar in the intermediate-level college French classroom". In: *Foreign Language Annals* 44.2 (2011), pp. 353–380.

[139] Qiuzhen Wang et al. "An eye-tracking study of website complexity from cognitive load perspective". In: *Decision support systems* 62 (2014), pp. 1–10.

[140] Aimee Weber, Kimberly Rufer-Bach, and Richard Platel. *Creating your world: The official guide to advanced content creation for Second Life*. John Wiley & Sons, 2007.

[141] Rustin Webster. "Declarative knowledge acquisition in immersive virtual learning environments". In: *Interactive Learning Environments* 24.6 (2016), pp. 1319–1333.

[142] Stefan M Wierda et al. "Pupil dilation deconvolution reveals the dynamics of attention at high temporal resolution". In: *Proceedings of the National Academy of Sciences* 109.22 (2012), pp. 8456–8460.

[143] John Williams and Patrick Rebuschat. *Implicit learning and second language acquisition*. Citeseer, 2016.

[144] Robert Wray et al. "Pedagogical experience manipulation for cultural learning". In: *Proceedings of the 2nd Workshop on Culturally Aware Tutoring Systems at the 14th International Conference on Artificial Intelligence in Education*. 2009, pp. 35–44.

[145] Chih-Hung Wu, Yi-Lin Tzeng, and Ray Yueh Min Huang. "A Conceptual Framework for Using the Affective Computing Techniques to Evaluate the Outcome of Digital Game-Based Learning". In: *Advanced Technologies, Embedded and Multimedia for Human-centric Computing*. Springer, 2014, pp. 189–196.

[146] M Wu. "Hofstede's cultural dimensions 30 years later: A study of Taiwan and the United States". In: *Intercultural Communication Studies* 15.1 (2006), p. 33.

[147] M Wu. "Hofstede's cultural dimensions 30 years later: A study of Taiwan and the United States". In: *Intercultural Communication Studies* 15.1 (2006), p. 33.

[148] Nick Yee. "The psychology of MMORPGs: Emotional investment, motivations, relationship formation, and problematic usage". In: *Avatars at work and play: Collaboration and interaction in shared virtual environments* 34 (2006), pp. 187–207.