University of Tennessee, Knoxville

TRACE: Tennessee Research and Creative Exchange

Doctoral Dissertations

Graduate School

5-2017

Exploring a Semi-Virtual Reality System Impacting Learning Curves of College Students

Hongbiao Yang
University of Tennessee, Knoxville, hyang22@vols.utk.edu

Follow this and additional works at: https://trace.tennessee.edu/utk_graddiss

Part of the Educational Psychology Commons, Industrial Engineering Commons, Other Education Commons, and the Other Operations Research, Systems Engineering and Industrial Engineering Commons

Recommended Citation

Yang, Hongbiao, "Exploring a Semi-Virtual Reality System Impacting Learning Curves of College Students. " PhD diss., University of Tennessee, 2017.
https://trace.tennessee.edu/utk_graddiss/4511

This Dissertation is brought to you for free and open access by the Graduate School at TRACE: Tennessee Research and Creative Exchange. It has been accepted for inclusion in Doctoral Dissertations by an authorized administrator of TRACE: Tennessee Research and Creative Exchange. For more information, please contact trace@utk.edu.

To the Graduate Council:

I am submitting herewith a dissertation written by Hongbiao Yang entitled "Exploring a Semi-Virtual Reality System Impacting Learning Curves of College Students." I have examined the final electronic copy of this dissertation for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, with a major in Industrial Engineering.

Rupy Sawhney, Major Professor

We have read this dissertation and recommend its acceptance:

James Simonton, Hamparsum Bozdogan, Eric R. Wade

Accepted for the Council:

Dixie L. Thompson

Vice Provost and Dean of the Graduate School

(Original signatures are on file with official student records.)

Exploring a Semi-Virtual Reality System Impacting Learning Curves of College Students

A Dissertation Presented for the

Doctor of Philosophy

Degree

The University of Tennessee, Knoxville

Hongbiao Yang

May 2017

Copyright © 2017 by Hongbiao Yang

All Rights Reserved

This dissertation is dedicated to my parents,Xuemei Tan and Changqing Yang, and my wife, Zhaoxia Zhao, for their love, support and encouragement.

Acknowledgements

Firstly, I would like to express my sincere gratitude to my advisor Prof. Rupy Sawhney for the continuous support of my Ph.D study and related research, for his patience, motivation, and immense knowledge. His guidance helped me in all the time of research and writing as well as data collection of this thesis. I could not have imagined having a better advisor and mentor for my Ph.D study.

Besides my advisor, I would like to thank the rest of my thesis committee: Prof. Hamparsum Bozdogan, Prof. Eric R. Wade, and Prof. James Simonton, for their insightful comments and encouragement, but also for the hard question which incented me to widen my research from various perspectives.

I thank my fellow labmates: Shuguang Ji, Gewei Zhang, Kaveri Thakur, Enrique Macias de Anda, Girish Upreti, Dhanush Agara Mallesh. Thank them for the stimulating discussions, for the sleepless nights we were working together before deadlines, and for all the fun we have had in the last four years. Also, I am grateful to Dr. Jerry Everett for enlightening me the first glance of research.

Last but not the least, I would like to thank my family: my parents and to my wife for supporting me spiritually throughout writing this dissertation and my my life in general.

Abstract

Virtual reality (VR) is a trending technology used in a broad range of fields including education and has become one of the most promising directions for educators. In this research, the investigation focuses on how the semi-immersive VR application can be used for educational purposes by exploring the VR factors and the interactions between these factors. A theoretical learning framework is also proposed to offer an explanation for the beneficial effects of education brought by VR at a high level.

This research consists of three parts. First, this research will introduce the development of Walk-in-Place Learning System (WIPLS), a semi-immersive VR system that is highly customizable and can be modified into different sub-VR systems that enable the tuning of various VR factors. Second, it will present the survey instrument obtained from previous literature related to educational VR systems. Two individual pilot studies will be conducted: 1) to verify the performance of the WIPLS, and 2) to validate the internal consistency of the survey instrument. Third, an empirical study will be conducted on a sample population to answer the research question, and to analyze the statistical results to validate the research model. Based on these statistical results, this research will propose conclusions and insights in how VR factors, as well as interactions, are affecting the learning outcome in an educational VR system, and provide guidance and suggestions for VR practitioners to design the development of VR systems.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Overview

Constructivist-based learning expounds that learning occurs as learners are actively involved in the process of gaining meaning and knowledge instead of passively receiving information (Glasersfeld, 1989). Studies also indicate that people with intrinsic motivation, who pursue the internal rewards of the learning process instead of the external stimulus, usually perform better in the learning outcome and retain the skill sets and knowledge more persistently. So, if a technology or system used for scholastics is based on the constructivist paradigm and can stimulate this intrinsic motivation, it may be believed with confidence that it will also facilitate the learning outcome. One such technique is Virtual Reality.

Virtual Reality (VR) is a computer user interface that involves the real-time simulation of an environment. It can provide the user with an immersive virtual environment which is believable and close to reality. VR has various promising characteristics, which makes it an ideal instrument for learning. VR can create an immersive and interactive environment that can make the learners actively involved instead of passively receiving the learning material, hence the constructivist teaching approach is supported. Also, with the 3-D visualization and enriched multimedia which are attractive to most of the learners, it may increase the participants' learning outcome by increasing their intrinsic motivation on the learning topics. The logic of how VR can impact

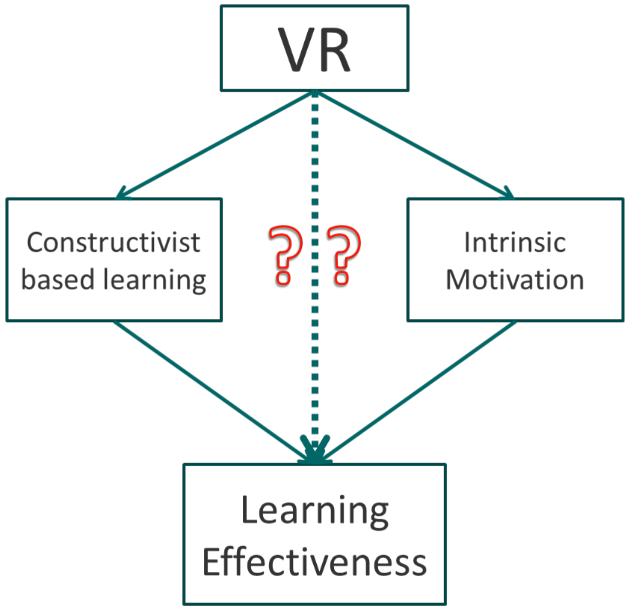

Figure 1.1: VR can impact learning effectiveness indirectly

the learning effectiveness is demonstrated in Figure 1.1. The dashed arrow implies the theory of VR impacting learning outcome is supported by literature, but the logical deduction behind it still requires additional exploration.

There are additional beneficial features of VR. The repetition capability within the VR technology can facilitate the participant's ability to practice specific tasks and scenarios, thus reinforcing the learning outcome. In regards to potential hazards experienced in other fields of the educational interface, especially to low-level learners, VR technology can be useful as it eliminates the risk of being exposed to the dangerous situation, therefore saving both financial and social cost. Traffic accidents that cause children fatalities and injuries are a good example where VR can play a significant role. Unintentional injuries are a leading cause of death and disabilities among children (Runyan et al., 2005), and one of the most common unintentional injuries are street accidents. Young children are more susceptible to street accidents because many complex perceptual and cognitive skills are required for a safe road crossing. These skills are developed over a child's age (Pitcairn and Edlmann, 2000). Traditional education techniques have limited effect in improving the road safety related abilities, as the adolescent trainees are too young to comprehend the traffic rules from textbooks. However, real roadside training brings potential risks to the trainees, thus it is considered unethical. In this regard, VR is an excellent tool for road safety training with child pedestrians.

1.2 Problem Statement

While many researchers have made the statement that VR-based learning system will achieve a better learning outcome because of the features above, few have explained how those features contribute to the improvement. Due to the relatively high cost of quantifying the impact of each factor of VR that contributes to the learning outcome, many researchers only compare the overall difference between the VR-based learning system and the control group. This usually confounds the actual impact of different variables (Cai et al., 2003; Coles et al., 2007; Ebner and Holzinger, 2007; Wrzesien and Raya, 2010). There is a lack of studies that decompose the VR learning system and analyze the contribution of each factor to the learning outcome. This research aims to take a closer look at the VR systems, analyze how VR technology promotes learning, and quantify how much each VR factor contributes to the learning outcome.

1.3 Research Motivation

While VR has bloomed in various fields in recent years, there is a lack of metrics that evaluate the actual contributions claimed by VR advocators. Many research experiments, both qualitative and quantitative, support the superiority of VR technology over traditional techniques. Most of those conclusions are drawn based on subjective and qualitative deduction, stating that since VR is highly motivating and attractive to users, apparently such increased attention will provide the users with more excitement and challenge, thus resulting in higher educational gains. This chain of logical deduction seems reasonable; however, few research conclusions attempt to quantify the causality behind this logic. Without quantifiable evidence, it is difficult to prioritize the factors that might be associated with VR technology. This makes the decisions to assign appropriate resources to a VR project difficult.

It is not easy to understand the factors and therefore makes it difficult to compare from one VR system to another. There is limited literature that decomposes the VR technology to the level of each factor of VR. The technology of VR is still in a trend with rapid development and iteration. This results in different commercial companies and academic institutions implementing their VR systems based on entirely different development instruments. This variation makes the comparison among VR systems almost impossible.

A customizable VR system is developed in this research. Such a VR system that can be modified with relatively low effort and cost can produce a series of VR systems that allows the ability to create similar VR systems while differing in only one factor at a time, to select certain pairs of VR systems that differ only by the factor of interest. This capability allows the explanation of any observed difference in the experimental outcome between these two VR systems based on specific factors.

1.4 Research Goals

The following are the specific goals of this research.

- *Development of a flexible VR system that allows the creation of customized sub-VR systems with varying levels of each factor.*

  This research will use the Unity3D gaming engine, a highly flexible gaming platform that supports sufficient degree of freedom on customizing the gaming to produce several heterogeneous sub-VR systems that originate from the same base model.

- *Develop key measures of learning outcome for a VR system.*

  This includes two parts. First, with selected VR factors, this research will develop a regression model that explores the quantitative relationship between the VR factors and the learning outcome. What's more, it will also propose a method to integrate theoretical learning frameworks including constructivist-based learning and intrinsic motivation. These two learning frameworks will be verified by checking the correlations between the VR factors and the critical components through statistical hypotheses.

- *Develop a survey instrument to measure learning outcome.*

  A survey instrument will be developed based on previous literature. The survey instrument will include survey items that measure the perceived learning effectiveness, satisfaction and the critical components of the theoretical learning frameworks. Reliability and validity of the survey instrument will also be evaluated.

- *Conduct an empirical study to verify the research concept in road safety.*

  To solve the regression model, this research will conduct an empirical study that makes use of the WIPLS. Design of Experiment will be used to provide guidance on how many sub-groups should be needed and how each sub-VR will be implemented from the WIPLS.

- *Analyze the statistical results* .

  After the empirical study is conducted, statistical methods will be performed on the experiment data and solve the regression model. The experiment data will also be used to conduct group mean comparison study to validate the theoretical learning frameworks.

1.5 Thesis Outline

The contribution of this research is demonstrated as follows. First, the VR system will be decomposed into its key factors. The list of relevant factors will be determined from a literature review. Next, this research will introduce the concept of constructivist-based learning (CBL) and intrinsic motivation (IM), as well as why those two approaches are beneficial to the learning outcome. The critical components that determine the learning outcome of those two approaches will also be discussed. Next, the effort will address how VR can support the CBL and IM by linking the VR factors to the components that are critical to those two theoretical learning frameworks. Lastly, after investigating a widespread range of VR applications, this research will contribute a case study that applies VR technology in road safety training and discuss the possibility of expanding the experiences and insights into other training and production-related areas.

This document is organized as follows: Chapter 2 will present a comprehensive literature review on each concept related to this research, including the CBL approach, the IM concepts, and the VR characteristics analysis. Also included in this chapter is a list of empirical studies on how VR might be beneficial in boosting the learning outcome. A pool of candidate VR factors will be extracted from these empirical studies. Chapter 3 will present the methodology of this research, which includes the conceptual framework, the research model, the development of the VR program, the survey instrument, the Design of Experiment, and the statistical method used in this research. Chapter 4 will focus on analyzing the collected data and interpreting the research results, using descriptive statistic, reliability measurement, validity measurement, factor analysis, ANOVA tests, and the hypotheses tests. Chapter 5 will focus on interpreting the statistical results reported in Chapter 4, and discuss the implications of the experiment finding and the insights for the VR practitioners. Chapter 6 will summarize this research, draw conclusions on the results of the research, and discuss the limitation along with future research following this study.

Chapter 2

Literature Review

2.1 Overview

The purpose of this chapter is to provide context for the research topic. The literature review will first define VR and how VR systems are categorized. Second, a special type of VR called Walking-in-Place Learning System (WIPLS) is introduced. WIPLS is used as the basis for the experiment in this research effort. Third, the CBL model that provides a theoretical basis on how VR can benefit learning outcome will be discussed. Fourth, IM theory is introduced, which is believed to be another positive stimulant to the student's learning outcome provided by the VR technology. Fifth, the learning effectiveness measurements used in this research is discussed. Sixth, a list of empirical studies on using VR for learning applications is presented.

2.2 Virtual Reality Categorization

Virtual reality (VR) is a computer user interface that involves real-time simulation of an environment, scenario, or activity that allows for user interaction via multiple sensory channels (Adamovich et al., 2009). VR has been widely used in various areas, mostly because it can create an immersive virtual environment that provides the users with realistic experiences which are otherwise costly or even impossible to obtain (Gutierrez et al., 2008).

Generally speaking, VR is divided into two categories based on the level of immersion the VR can support (Adamovich et al., 2009). The first type, also known as immersive VR, usually comes with the head mounted display (HMD) or wall-sized screen surrounding the users that provide stereoscopic visualization. This allows the users to feel the presence of being in a special environment. The other category, non-immersive VR, usually uses commodity-level hardware for visualization, like a normal monitor or TV screen. Technically speaking, the users in this category of VR are not immersed in the virtual environment; instead, the experience is similar to observing through a window (Lee, 2011). Besides the visualization perspective, there is another dimension to consider when evaluating the immersion level. This immersion level is associated with the naturalness of the interaction, or transduction (Winn et al., 1993), between users and VR. An immersive VR usually interacts with the users in natural semantics. For example, when manipulating a 3D virtual object in the virtual world, moving and rotating it using hands and fingers are considered natural and immersive. However, if the users do so by dragging and clicking the mouse, it is less natural and provides limited immersion, since it contradicts with the users' intuition and experience gained from the real world.

With the two dimensions in the VR categorization, all VR can be divided into four quadrants, as displayed in Figure 2.1.

A VR system with both stereoscopic 3-D display and natural controllability is easily defined as full immersive VR. Similarly, if a VR system only uses normal screen for display and standard input devices like a keyboard and mouse for controlling, it is defined as a non-immersive VR, which can also be called desktop VR. The VR systems that fall in the other two quadrants cannot be easily defined as either immersive VR or non-immersive VR. It is best considered the mediation category of these two groups: the semi-immersive VR. This category has a broad range of instances, for example, wearing an HMD to play a traditional commercial console game using a standard game controller, or using a motion sensor input device to play games on a laptop.

When the user experience is considered, there is almost no doubt that immersive VR will outperform the non-immersive VR, since it provides more immersion, interaction and visualization. However, this does not imply that non-immersive VR is without benefit. Non-immersive VR system is more affordable and is an excellent option when budgets are an issue.

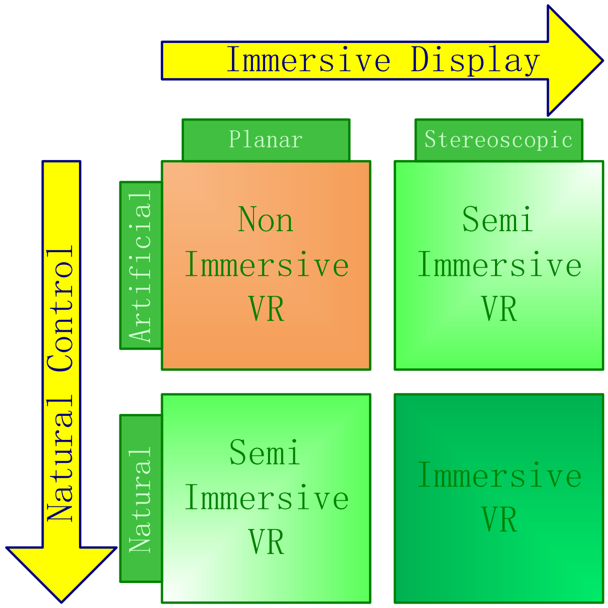

Figure 2.1: Categories of VR according to different perspectives

The full immersive 3-D visualization can be an astonishing experience for some users, while it can also cause motion sickness (Bangay and Preston, 1998) to another group of users who are uncomfortable with the 3-D virtual environment or have a lower spatial ability.

As mentioned before that the semi-immersive VR is a compromise of the two categories, it can find a VR that reaches a balance based on needs of training, budgets, target population, etc. The semi-immersive VR systems are pragmatic solutions for VR applications, which provide a greater amount of immersion and interaction than the desktop VR, while at the same time remaining at a relatively low cost.

2.3 Walking in Place - a Subcategory of VR

Walking-in-Place (WIP) system is a type of VR system that allows locomotion in the virtual environment within a small physical space. The goal of the WIP system is to allow the user to move in a virtual environment in ways similar to walking in the physical environment (Templeman et al., 1999). Some researchers have taken an interest in the design of advanced WIP systems.

For example, Low-Latency, Continuous-Motion (LLCM) WIP (Feasel et al., 2008) is a high-performance WIP system. The developers used sensors to collect chest orientation and heel speed data; they then converted these data into direction and motion in the virtual world.

Various implementations of WIP systems (Bruno et al., 2013; Feasel et al., 2008; Kim et al., 2012; Wendt et al., 2010; Whitton and Peck, 2013; Williams et al., 2011; Zielinski et al., 2011), all of them suffer from common problems like latency (Yan et al., 2004), jerkiness (Multon and Olivier, 2013), and user burden (Psotka, 1995). Starting/stopping latency is a fundamental problem for accurate simulation of realistic forward motion(Yan et al., 2004). Too much latency causes cyber sickness(Sibert et al., 2004). Latency also results in unrealistic virtual collisions (Usoh et al., 1999) during walking, detracting from the immersive nature of the virtual interaction. Another issue is the jerkiness between adjacent steps. Jerkiness is a term in motion pictures that refers to a series of distinct snapshots instead of smooth and continuous motion and is usually caused by dropped frames (Huynh-Thu and Ghanbari, 2006). Jerkiness can result in the non-fluent and non-smooth presentation of video (Borer, 2010) that annoys video viewers (Lin and Jay Kuo, 2011) and detracts from the experience. In WIP systems, jerkiness can reduce the feel of realism and immersion in the virtual environment (Stakem et al., 2007). In addition to these two problems, device calibration, and user burden is also considered important factors that impact the WIP system.

2.4 Learning Outcome

Learning outcome is defined as "not only the knowledge leading to understanding but also abilities, habits of mind, ways of thinking, attitudes, values and other dispositions." (Maki, 2012). Objective measurements and subjective measurements including the improved academic achievements, self-esteem, quality of interpersonal interaction and student attitudes are also considered to be affecting the learning outcome (Johnson et al., 1998; Prince, 2004; Springer et al., 1999). In this study, learning outcome refers to a combination of perceived learning effectiveness and the satisfaction towards the learning tool.

The perceived learning effectiveness measures the amount of information participants thought he/she learned effectively through a learning activity. It has been widely used in numerous studies as a measurement instrument (BENBUNAN-FICH and HILTZ, 2003; Lee, 2011; Marks et al., 2005). The satisfaction is a more subjective measurement instrument that measures how the participant is satisfied with the learning method provided.

2.5 Using Constructivism as an Approach

2.5.1 Constructivism Definition

Constructivism is about how people learn, with the belief that learners construct their own knowledge interactively based on what they already know, instead of receiving knowledge from the teachers passively following a fixed structure (Brooks and Brooks, 1999). It emphasizes stimulating the learners to engage in the process of learning actively (Felix, 2002).

There are several differences between the traditional classroom and the constructivist one, as displayed in Table 2.1 (Brooks and Brooks, 1999).

2.5.2 Advantages of using Constructivism in Learning

The greatest advantage of constructivist learning is that the learners do not need to memorize separated, isolated parts of the problem to pass quizzes and tests; instead, they are encouraged to foster new skills and knowledge based on what the learners already know (Lefoe, 1998). The constructivist learning approach believes that knowledge is constructed through the participation of certain experiences, and it provides the realistic experience to the learners that enables them to construct their knowledge and skill through the process of solving an authentic problem (Lainema and Makkonen, 2003). In traditional objective learning, learners usually receive knowledge from highly abstracted theories and concepts. Thus there is an additional transfer process from abstraction to a particular skill required in a practical problem.

Table 2.1: Comparison between a traditional and a constructivist classroom (Brooks and Brooks, 1999)

Traditional Classroom	**Constructivist Classroom**
Curriculum begins with the parts of the whole. Emphasizes basic skills.	Curriculum emphasizes big concepts, beginning with the whole and expanding to include the parts.
Strict adherence to fixed curriculum is highly valued.	Pursuit of student questions and interests is valued.
Materials are primarily textbooks and workbooks.	Materials include primary sources of material and manipulative materials.
Learning is based on repetition.	Learning is interactive, building on what the student already knows.
Teachers disseminate information to students; students are recipients of knowledge.	Teachers have a dialogue with students, helping students construct their own knowledge.
Teacher's role is directive, rooted in authority.	Teacher's role is interactive, rooted in negotiation.
Assessment is through testing, correct answers.	Assessment includes student works, observations, and points of view, as well as tests. Process is as important as product.
Knowledge is seen as inert.	Knowledge is seen as dynamic, ever changing with our experiences.
Students work primarily alone.	Students work primarily in groups.

2.5.3 Strength of using VR to Support Constructivist Learning

VR technology is capable of supporting the constructivist learning because it can provide features like interaction, immersion, visualization, and natural semantics, which can be used as important factors required for constructivist learning (Winn et al., 1993). Being immersed in a virtual environment that is realistic and interactive, the learners can intuitively apply the prior knowledge and experience in the new tasks to solve certain problems in an authentic form. There will be no more necessity to invest cognitive effort to comprehend the narrative problem in the form of text and static images which are in the abstract form of the knowledge. Through the process of solving the problem in the virtual environment, knowledge and skills are constructed by the learners themselves, not transferred from the outside environment by memorization (Felix, 2002).

2.6 Intrinsic Motivation

Extrinsic and intrinsic motivations are two types of human motivation that drive people to perform certain actions. Extrinsic motivation involves doing something for the external reward, like money, praise, or anything that is tangible. On the contrary is intrinsic motivation, which refers to the fact of doing an activity for itself, and seeking internal reward like pleasure and satisfaction that are derived from participation (Ryan and Deci, 2000).

Most researchers believe that extrinsic motivation can stimulate students to gain initial interest and engagement in some situations, as this stimulation usually fades quickly and will undermine any intrinsic motivation the students already have, if administered improperly (Lepper et al., 1973). Intrinsic motivation, on the other hand, is driven by interest, which will usually be long lasting and more creative and productive since the students gain pleasure and motivation from the task itself (Coon and Mitterer, 2012). A positive correlation between intrinsic motivation and academic achievement has already been found in educational studies (Pintrich and de Groot, 1990).

There are three types of intrinsic motivations, according to (Vallerand et al., 1992). These are: intrinsic motivation to know, intrinsic motivation toward accomplishments, and intrinsic motivation to experience simulation. The first intrinsic motivation (IM to know) is the desire and curiosity to explore and understand something new. The second intrinsic motivation (IM to accomplish things) results from the pursuit of the pleasure of satisfaction when attempting to accomplish something or master some skills. The last type of intrinsic motivation (IM to experience simulation) is operative when someone engages in some activity to experience the simulating sensation derived from the engagement of the activity.

VR can support intrinsic motivation since VR satisfies several elements that can foster intrinsic motivation such as choice, control, collaboration, challenge, and achievement (Malone and Lepper, 1987). More importantly, VR is excellent at providing immersion; through which the learners can have the opportunity to feel the presence of being in another environment, thus, engaging in those immersive activities and gain IM to experience simulation (Huang et al., 2013). According to (Winn et al., 1993), VR can also provide a first person, non-symbolic experience for students, which can motivate a large number of students who do not master the symbol systems of the disciplines in their study.

2.7 Current VR for Learning Applications

VR has been widely used in various fields including military (Piekarski et al., 1999), medical (Mann et al., 2002; Satava, 1995), rehabilitation (Kim et al., 2009; Mirelman et al., 2010; Wade and Winstein, 2011) , education (Coller and Scott, 2009; Pan et al., 2006; Vera et al., 2005), and so on. In manufacturing field, VR has also been used in various aspects like assembly design (Jayaram et al., 1997), prototyping (Choi and Cheung, 2008), and employee training (Olive et al., 2006; Mujber et al., 2004) .

Education has been a major player in VR applications since its invention due to its unique characteristics of creating immersive virtual worlds and attracting subjects. A key target group is children who do not have a well-developed cognitive ability in understanding abstract concepts. VR technology has been applied in education field by many researchers on various subjects. These researchers reported different VR factors in their empirical studies listed in Table 2.2. Based on these empirical studies, a pool of VR factors was identified. The VR factors in this pool were reported to be beneficial in affecting the learning outcome of VR. Thus they can also be considered as candidate VR factors in this research. According to Table 2.2, there are a vast number of empirical studies that tried to examine the benefits of applying VR in various fields of education. However, none of these empirical studies explored how the learning outcome is affected by the VR factors systematically, which is what will be addressed in this research.

The easiest way of applying VR in education is directly combining the game characteristics and educational content together. In this type of VR, the students are required to complete some learning activities before they can gain access to the entertainment contents, whether those entertainment contents are relevant to the learning activities or not. This type of VR is called "carrot and stick" (Charsky, 2010). For example, (Holzinger et al., 2006) use a quiz show game called TRIANGLE as an interactive multimedia learning object to teaching students mathematics. The students' tasks in TRIANGLE are to achieve as high a score as possible by answering ten questions. Similar to the TRIANGLE, (Virvou et al., 2005, 2002) used a VR-based educational game called VR-ENGAGE to teach the students geography. In this educational VR game, the students are required to navigate through a virtual world and continue their way by answering questions posed by a virtual character. The authors compare the learning effectiveness regarding testing scores in the quiz between the VR-based educational game and the traditional educational software. Although those two aforementioned VR-based educational games are proven to have a positive impact on the learning effectiveness, specifically in some types of learning that require fact rote and memorization, in some other types of learning where complex and flexible understanding of knowledge and application of skills are called for, this "carrot and stick" strategy is insufficient (Charsky, 2010). According to the criterion brought up by (Winn et al., 1993), VR of such types can only be deemed as "third person, symbolic" systems, which requires deliberate reflection between the abstract symbol system and the actual learning experience.

A better form of VR system for education would be those that support free discovery and navigation. Such VR systems may provide a more "seamless integration" between VR and the learning content, and affords a "first-person non-symbolic experience" (Winn et al., 1993). It is also considered as a more natural way of interaction (Bricken, 1991). In the research conducted by (Wrzesien and Raya, 2010), a serious virtual world called E-Junior is developed which allows the

Table 2.2: Empirical studies applying VR in education field

Year & Author	**Factors**	**Application**	**Comments**
Charsky, 2010	Entertainment	Education	Carrot and stick paradigm
Holzinger, Pichler, & Maurer, 2006	Interaction	Education, Math	VR name: Triangle
Winn et al., 1993	Immersion, natural semantics	Education	First person, non-symbolic
Bricken, 1991	Natural Semantics	Education	Discusses educations using VR in general
Wrzesien & Raya, 2010	Interaction, visualization	natural science, geography, and ecology	VR name: E-Junior project
Cai et al., 2003	Interaction, free navigation	Biomedical domain	
Meluso, Zheng, Spires, & Lester, 2012	Immersion, interaction, collaboration	Science related education	VR name: CRYSTAL ISLAND
Erhel & Jamet, 2013)	Entertainment	Medical related assessment	VR name: Digital game-based learning
(Moreno & Mayer, 2005	Interactivity, reflection, feedback, guidance	Science learning	VR name: Agent-based multimedia games
Mayer, Mautone, & Prothero, 2002	Visualization	Geology task training	VR name: The Profile Game

Table 2.2: continued

Year & Author	Factors	Application	Comments
Pausch, Proffitt, & Williams, 1997	Immersion	Performing searching tasks	VR name: CAVE
Vora et al., 2002	Immersion, presence	Aircraft inspection training	Immersive tendencies questionnaire (ITQ) and presence questionnaire (PQ) were used for assessment
Bangay & Preston, 1998	Excitement of the experience, comfort of peripherals and environment during the experience, quality of the sound and images	Pure experience theme park	Swimming with dolphins and virtual roller coaster were used as experiments
Holzinger et al., 2006	Attraction, Fun, Challenge, Fantasy, Curiosity, Interaction, graphics	Mathematics curriculum study	VR name: TRIANGLE
Virvou et al., 2005	Free Navigation, Interaction	Knowledge of Geography	VR-ENGAGE is the VR system. Former VR gaming experience was studied as a factor in this research

Table 2.2: continued

Year & Author	Factors	Application	Comments
Blackledge & Barrett, 2012	Safety	Engineering education, training of safety-related knowledge and skills	VR name: Virtual Electrical Services
Coles et al., 2007	Safety, repetition	Street safety and fire safety education	Children with FAS(fetal alcohol syndrome) are used as subjects in this research
Simpson, Johnston, & Richardson, 2003	Children pedestrian safety	Demographic factors	Use road crossing related counts and timing measures as response variables
Meir, Parmet, & Oron-Gilad, 2013	Pedestrian safety	Road-crossing scenarios related factors	Different road-crossing scenarios are studied in detail in this research
Schwebel, Gaines, & Severson, 2008	Children pedestrian safety	Demographic factors, children temperament	Parents are invited to participate in the experiment together with the children

students to freely navigate and explore in the virtual aquatic world using a navigational input device to learn nature, science, and ecology. The author evaluated the learning effectiveness of the E-Junior application by comparing the virtual group with a traditional class. The knowledge test in the pretest and posttest for both groups proves that the serious virtual world groups do not present statistically significant difference from the traditional class, while the students in the virtual group reported more enjoyment and engagement as well as more intention to participate. The biological virtual environment brought up by (Cai et al., 2003) is a game based problem-solving environment that allows the users to explore biological interactions. This problem-solving environment provides the students with navigation on atomic to macroscopic scales, role-play, and networked collaboration. A case-study is presented in a group of young children with no background where certain quantified variables are analyzed to measure how much those children have learned through playing a game-based problem-solving environment. In addition, some learning assessment questions are also asked. The results of the experiments show that this game greatly inspired users both in concept learning and entertainment. The study conducted by (Meluso et al., 2012) investigated the effects of collaborative and single game player conditions on science content learning and science self-efficacy. The authors used an online computer game called CRYSTAL ISLAND, which consists of an immersive 3-D intelligent learning environment with a cast of characters within a story world. Through navigating in the virtual world and interacting with the virtual characters, the students will have the opportunity to learn about science-related concepts. Results show that there was no significant difference between the two playing conditions, while a significant increase was found in science content learning and self-efficacy in the posttest assessment compared to the pretest assessment when collapsing those conditions.

However, not everyone agrees that more degree of freedom is always better. (Mayer, 2004) criticize that pure discovery may sometimes distract the students from the to-be-learned material, and virtual discovery under some instructional guidance is more effective in helping students learn and transfer. An experiment conducted by (Erhel and Jamet, 2013) tries to study the conditions under which digital game-based learning (DGBL) is most effective through analyzing the effects of two different types of instructions: learning instructions and entertainment instructions. In one of the experiments conducted in this research, the participants are interacting with a multimedia learning environment called ASTRA, which takes the form of a simulated living room where a female pedagogical agent stands next to a TV screen and provides the oral information and instructions to the participants. The results of this experiment reveal that comprehension scores were significantly higher in the learning instructions condition than in the entertainment instruction conditions. This supports the arguments that the game-based environment without any instructional guidance may not achieve the positive learning effectiveness as expected. The study conducted by (Moreno and Mayer, 2005) investigated whether the guidance and reflection would facilitate science learning in an interactive multimedia game. 105 undergraduate freshmen were recruited in the experiment. They were divided into four treatment groups with two treatment factors tested; the first treatment factor is whether or not they were asked to explain the answer, and the other one is whether or not they received an explanation of the answer after being told whether they were correct. The results of the experiment show that guidance is a significant effect in an agent-based multimedia game. In conclusion, the authors demonstrate that designers of agent-based games should incorporate structured guidance rather than rely solely on pure discovery. Another study that supports the guided discovery in a virtual environment is conducted by (Mayer et al., 2002)). In this paper, the authors examine what type of guidance will be most beneficial to help students solve problems within a multimedia simulation environment. They use Profile Game, which is designed to represent authentic tasks that the geographers perform in their scientific work based on visual data. Participants are college students divided into subgroups. Four different guidance conditions are tested in the experiments, which are: 1) illustration of possible geological features, 2) verbal descriptions of how to solve problems, 3) both illustration and verbal description, and 4) the control condition where no guidance is given. Through a series of experiments, the authors conclude that the best performing group is the 3rd group that received most guidance in the virtual environment. This result is consistent with the research showing that guided discovery is a better solution compared with pure discovery with no guidance.

Knowing that VR can boost the learning effectiveness and satisfaction is not enough; it is still necessary to look into the VR system and analyze which factors actually improve the learning outcome. Several studies have been conducted on whether or not those factors would play a major role in the improvement of learning effectiveness of VR. Immersion is undoubtedly the first factor that one would think of when talking about VR. The research conducted by (Pausch et al., 1997) compared the performance of carrying out a search task between the VR users with a VR interface and the desktop users with a stationary monitor and a hand-based input device. The authors found that VR users were substantially better at determining when they had searched the entire room than the desktop users. From the experiment results, the authors concluded that VR could improve user performance via immersion. In the research of (Vora et al., 2002), immersion was also proved to be a significant factor in VR. This research measured the degree of immersion and presence felt by subjects in a virtual environment simulator for aircraft inspection training. The authors tried to explore subjective presence as they believed it might affect the task performance. The results of the experiments indicated that the VR system in this research demonstrated high scores on most of the aspects of the presence issue, stating that it can suitably mimic the real world environment. Also, in the comparison between the VR system and the PC-based simulator system with no immersion, the VR system also proved to be better and more favored. However, immersion is too big a topic to be simply considered as one single variable. (Bangay and Preston, 1998) tried to decompose this variable, and identified the factors that may affect or be affected by the degree of immersion in a VR system. Two virtual environments are used on participants at a school science festival to collect heart rate data, head movement data, and feedback from questionnaires. These two virtual environments are: “swimming with dolphins” and “virtual roller coaster”. From the results of the experiments, the author demonstrates that the factors that influence the effectiveness of immersion in a VR environment are: excitement of the experience, comfort of peripherals and the environment during the experience, quality of the sound and images, and participants’ age. The author also found some factors that show a dependence on the degree of immersion, which are: simulator sickness, control, excitement of the experience, and desire to repeat the experience. Beside immersion, there are still more factors that are worth analyzing. In the experiment conducted by (Holzinger et al., 2006), three main factors of VR regarding learning are tested, which are: motivation, incidental learning, and a concept of personal responsibility named Tamagotchi effect. Questionnaires, objective data from user tracking log-file, and questions are used to test those three concepts. According to the authors, the results of the experiments showed significant differences between the experiment group and the control group on motivation and incidental learning, while no significant difference is found on the Tamagotchi effect, which means the presence or absence of the avatar in the VR program did not have any observable influence.

Besides the factors of the VR system, some demographic information factors of the participants are also worth exploring, like former academic performance and gaming experience. According to (Virvou et al., 2005), former performance may be an important factor since students with different former performance may also gain different learning outcomes from the VR-based learning. Experiments are conducted on the students both in general and in subgroups, where subjects are divided into three subgroups based on their former performance. The results of this study show that there are significant improvements in the posttest for the whole population in general. For the experiments conducted in subgroups, the students who formerly performed poorly and averagely show significant improvements in the posttest, while there is no significant difference in the posttest for the subgroup of students whose former performance was good. This study provided the insights that former performance may be a factor that affects how much a student can benefit from the VR-based learning approach. Another experiment also conducted by (Virvou and Katsionis, 2008) addressed the issue of usability and likeability of a virtual reality game for students' geography teaching affected by former gaming experience. This time, the authors divided the students into three groups according to their level of game play expertise, which are: novice, intermediate, and expert. For the usability issue, the authors took user interface acquaintance, navigational effort and VR environment distractions as three features for interpretation. The authors concluded that the usability problem does exist to some extent in VR environment, and the novice users are most easily to be affected. For the likeability issue, the users are asked to play a VR education game versus non-game educational software in the classroom setting for comparison, as well as play a VR education game versus commercial non-educational game at home for comparison. The results show that the students are harder to satisfy playing the game at home than in the classroom.

In addition to those quantifiable variables, there are still some advantages of VR that are difficult to quantify, for example, safety issues. (Blackledge and Barrett, 2012) uses a desktop VR named Virtual Electrical Services that can provide an appealing training and design environment and allows the users to operate in a safe environment and may potentially reduce the training costs and enhance electrical safety. In the case study conducted by the authors, several measurement items are collected through questionnaires, including VR features, usability, learning experience, and VR model measurement outcomes. After the experiments, a group discussion is used for additional qualitative feedback. In the end, this research concluded that the developed prototype has the potential to increase understanding of issues related to electrical safety and could potentially help cut down on accidents and fatalities related to electrical shock and electrocution, and users were receptive to using VR as the learning and design tools. VR can also be used for training of safety-related knowledge and skills on particular groups of population without bringing the participants into risky situations. The research conducted by (Coles et al., 2007) studies whether or not the children with fetal alcohol syndrome (FAS) can learn fire and street safety knowledge and skills through a computer game that employed "virtual world". 32 children aged 4-10 with FAS were recruited to participate in this game. After playing the game repeatedly, the children were retested both verbally and behaviorally and were given a follow-up test one week later. The authors demonstrated that after the experiment, the children showed significantly greater knowledge gains in both the verbal and behavioral test, and the authors concluded that the computer game with multisensory learning experience is a highly effective method for teaching high-risk children safety skills. Besides the safety issues, the authors also propose that VR technology can afford repetitions of learning activities to the participants until the mastery of the skills is achieved, which is an extraordinary virtue, especially for the population with limited cognition, since they may require more repetition than people with normal cognitive ability, and such repetition may be aversive or boring for the human instructors.

When it comes to the scope of pedestrian safety in the children and young adult's population group, the safety feature becomes a crucial factor of VR. (Simpson et al., 2003) investigated the road crossing behavior of children and young adults using a VR system and head mounted display. Two sets of experiments were designed: uniform speed and uniform distance, with age group, gender, and trial type as the independent variables. The response measurements include counts of unsafe crossing and cautious crossing as well as timing measures. The authors concluded that VR is advantageous since it is more real than the "shouting task", and less dangerous than the road side approach. (Meir et al., 2013) used a Dome-Projection Environment to simulate various road crossing scenarios, including zebra crossing, restricted field of view, and moving vehicles. The authors divided the participants into four groups according to their age. The experiment results indicated that the youngest group of children and the adult group were the most performant groups. The authors interpreted the results with different reasons showing that the adults were experienced and making comprehensive decisions, while the young children achieved good performance only because they were less aware of the potential hazards. (Schwebel et al., 2008) conducted experiments that compared different methods of road safety interventions including VR, shout out technique, two steps technique, and real road-crossing (only for the adult group) among the children group and adults group. Continuous variables (gap size available, average wait time, and average start delay) and discrete variables (counts of error and close calls) are included as response variables in this experiment. The authors concluded that VR could be considered as an appropriate methodology for both etiological research on the causes of pediatric pedestrian injuries, and for intervention research designed to study virtual reality as a tool to train children in pedestrian safety. This research not only included demographic factors like age, gender, race, and socioeconomic status as the independent factors, but also contributed a new factor called temperament of the children. None of these studies considered taking the VR features as the independent factors when conducting the pedestrian safety related research.

2.8 Summary

This chapter provides an overview of VR technology and the possibility of using VR as a teaching instrument to enhance learning outcome through CBL approach and by increasing the intrinsic motivation. To exert the VR technology as an effective learning instrument, merely combining the game characteristics of VR and the learning content is not enough. Instead, the educational VR systems that seamlessly integrate the VR features and the learning content are better choices.

Also, to find out how VR can boost learning outcome, it is necessary to look into the VR systems and determine which factors are playing the significant roles. From the literature review, a list of empirical studies was investigated and the corresponding results were displayed in Table 2.2. A pool of candidate VR factors as well as demographic factors was generated for further research.

Also, it has been found that none of these empirical studies tried exploring the learning outcome affected by individual VR factors systematically. This research is proposing to solve this issue.

In chapter 3, a research model that incorporates those proposed VR factors will be formulated, and corresponding hypotheses will also be generated to validate this model.

Chapter 3

Methodology

3.1 Overview

This chapter presents the theoretical foundation of this research. First, this research discusses the conceptual framework, including the definition of the learning outcome and descriptions of the selected VR factors from literature review and of the theoretical learning frameworks. Next, the VR program developed for this research, which is called WIPLS, is introduced. A pilot study is also discussed with the purpose of evaluating the performance and characteristics of WIPLS. After that, the Design of Experiment (DOE) is discussed, with the design choice and explanation on which pattern of DOE is applied in this research. Next, this research presents the survey instrument extracted from literature review. The survey instrument was used in the case study to evaluate the learning outcome from the participants. Since some changes were made to the survey instrument to better fit the scope of this research, another pilot study is conducted to evaluate the reliability and validity of the modified version of the survey instrument. Next, this research describs an empirical case study, including the participants' information, the implemented sub-VR systems, and the experimental procedure. Finally, the statistical methods for analyzing the experiment data are discussed. The complete research structure is demonstrated in Figure 3.1.

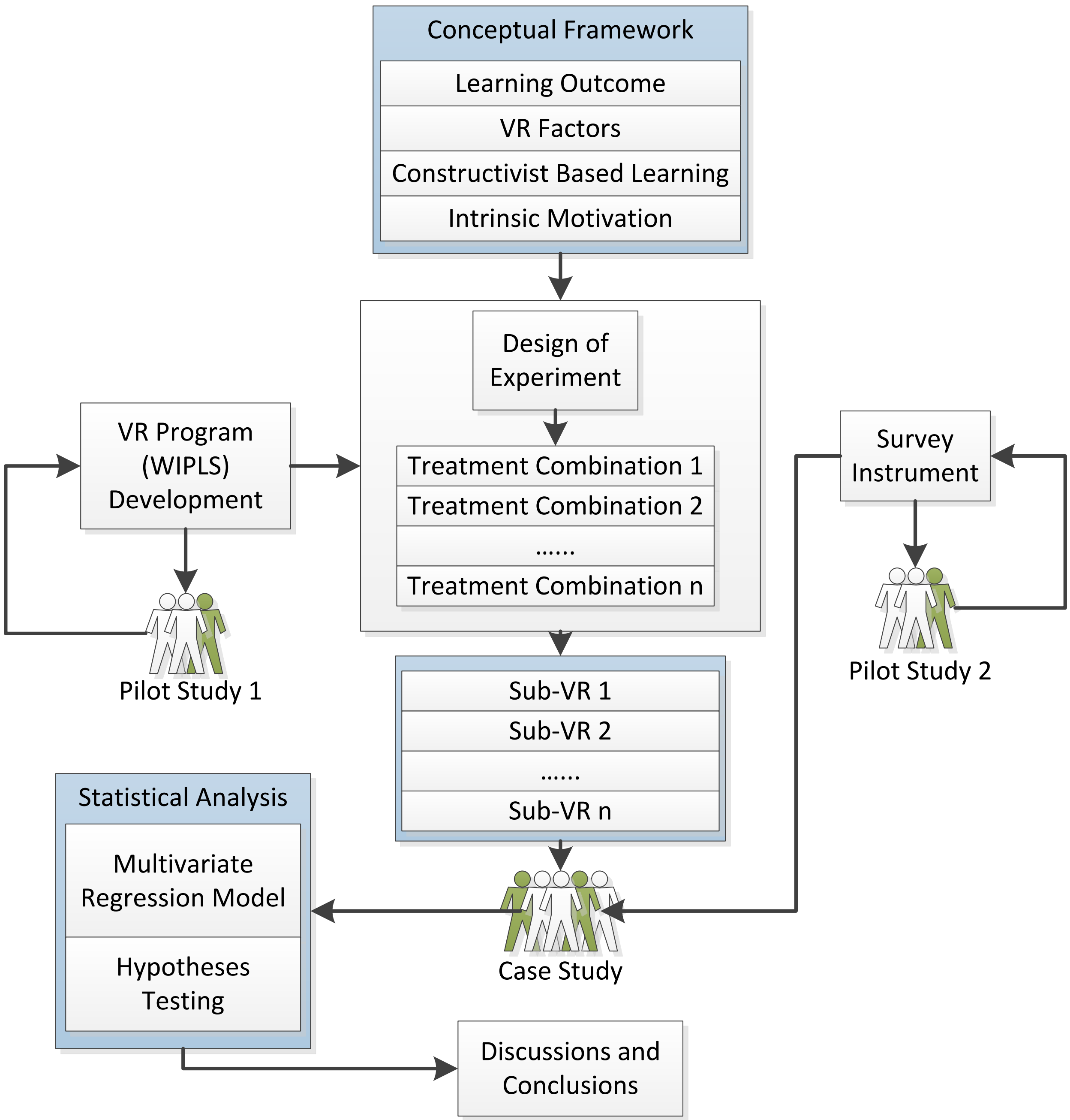

Figure 3.1: Structure of the research contribution.

3.2 Conceptual Framework

In this section, the conceptual framework of this research is discussed. First, the learning outcome is defined. Next, a short list of VR factors are selected from the candidate VR factor pool for further research. Finally, the CBL and the IM are analyzed in detail and decomposed into critical components that can be evaluated in further research.

3.2.1 Learning Outcome

Learning outcome in this research is interpreted as the combination of perceived learning effectiveness and the satisfaction towards the learning tool.

'Perceived learning effectiveness' is an objective metric that measures the amount of information participants thought they learned effectively through a learning activity. 'Satisfaction' is a subjective measure of the level of satisfaction of the participant with the provided learning method. These two measurement instruments are evaluated using the survey instruments, which are discussed in detail in Section 3.5.

3.2.2 VR Factors that Affect Learning Outcome

According to the literature review listed in Table 2.2, several factors can impact the outcome of the learning process. The list of factors includes:

- Free navigation
- Visualization
- Natural semantics
- Autonomy
- Presence
- Interaction
- Exploration
- Collaboration
- Immersion

It was unnecessary and impractical to study every single VR factor obtained from the literature review. Instead, only a short list of VR factors were selected. The criteria for selecting VR factors for research are as below:

- *Popularity*

  Only the factors that are mentioned most in literature are taken into consideration.

- *Independence*

  The selected factors need to be independent from each other. Factors that are highly correlated can be combined.

- *Measurability* The selected factors need to be objectively measured in the experiments. Factors that heavily rely on subjective perception are skipped in this research.

- *Significance*

  VR factors that were reported to be significant in previous empirical studies are selected. Factors that were proven to be insignificant are ignored in this research.

- *Practicability*

  We are planning to manipulate the levels of VR factors and compare the corresponding learning outcome to study the effects of those VR factors. If a VR factor cannot be implemented in the customizable VR system, it is excluded from the study.

The complete list of candidate VR factors and the selected ones are presented in Table 3.1. For each VR factor that is not selected, the specific violated criteria are also stated.

The short list of VR factors and the definitions are presented as below:

- *Visualization*

  Visualization in this research refers to characteristics of VR that affect users' visual sensation, including the quality of graphics, stereoscopic effects, color scheme, display resolution, etc. (Wrzesien and Raya, 2010).

Table 3.1: Candidate VR factors selected according to the selection criteria

	Selected	Criterion violated	Literature source
Free Navigation	N	Popularity	Cai et al., 2003;Virvou et al., 2005
Visualization	Y	-	Wrzesien & Raya, 2010; Mayer, Mautone, & Prothero, 2002
Natural Semantics	Y	-	Winn et al., 1993; Bricken, 1991
Autonomy	N	Independence	Ryan, Rigby, & Przybylski, 2006
Presence	N	Independence	Vora et al., 2002
Interaction	Y	-	Holzinger, Pichler, & Maurer, 2006; Wrzesien & Raya, 2010;Cai et al., 2003;Meluso, Zheng, Spires, & Lester, 2012;Holzinger et al., 2006;Virvou et al., 2005
Exploration	N	Measurability	Satava, 1995
Collaboration	N	Practicability	Meluso, Zheng, Spires, & Lester, 2012
Immersion	Y	-	Winn et al., 1993;Meluso, Zheng, Spires, & Lester, 2012;Pausch, Proffitt, & Williams, 1997;Vora et al., 2002
Sound quality	N	Popularity	Bangay & Preston, 1998
Image quality	N	Popularity	Bangay & Preston, 1998
Tamagotchi-effect	N	Significance	Holzinger et al., 2006
Entertainment	N	Practicability	Charsky, 2010

- *Natural semantics*

  Natural semantics is defined as the manner of behavior that is intuitive and natural, with the objective to minimize the burden of learning new knowledge and make use of what the users already know (Winn et al., 1993).

- *Interaction*

  Interaction in the domain of VR is the pattern of sending commands and directions to the VR system as the sender, as well as receiving feedback from the VR as the receiver (Nalbant and Bostan, 2006).

- *Immersion*

  Immersion of VR is defined as the sense of being in an environment while the user is physically in another environment (Pausch et al., 1997).

3.2.3 Critical Components that Affect Learning Outcome

As per the literature review, CBL is able to positively affect the learning outcome, since there are several critical components of the CBL that are linked to the learning outcome. Those critical components include: Active Learning, Interactive Learning, and Authentic Problem. Similarly, the increased IM also affects the learning outcome not directly but through critical components, which include: Control, Challenge, and Experience. Table 3.2 demonstrates the critical components of CBL and IM as well as the corresponding literature sources. The descriptions of those critical components are as below:

- ▶ *Active learning*

  CBL allows students to construct knowledge based on what they already know, instead of passively receiving didactic instructions from the teachers. The traditional teacher-to-student way of transferring knowledge is changing to one in which students actively seek knowledge on their own, with the eagerness to explore the subject and find answers to the questions raised by the teacher. The teacher is not an irreplaceable role in this education paradigm, but

Table 3.2: Critical Components of CBL and IM and literature sources

Theoretical learning frameworks	Critical components	Literature sources
Constructivist-based learning	Active learning	Lee, 2011, Grabinger & Dunlap, 1995
	Interactive learning	Harper & Hedberg, 1997; Huang, Backman, Chang, Backman, & McGuire, 2013
	Authentic problem	Chuang & Tsai, 2005, Mayer, Mautone, & Prothero, 2002
Intrinsic Motivation	Control	Dickey, 2006; Waterman et al., 2003
	Challenge	Dickey, 2006, Ryan & Deci, 2000
	Experience	Ryan & Deci, 2000; Huang, Backman, Chang, Backman, & McGuire, 2013;Waterman et al., 2003

more like an assisting coordinator who can provide help and guidance when students face difficulties and are looking for help. This component is crucial to the CBL mode, through which knowledge can be constructed by the students effectively, and the forgetting curve is believed to be far more flat than the traditional learning mode.

▶ *Interactive learning*

There are two types of interaction that are thought to be beneficial in learning activities according to CBL approach. The first category is the interaction among learners and instructors. Constructivists believe that learning occurs not in isolation from others, but through interaction among participants (Huang, 2002). With synchronous and asynchronous communication tools, such as group chat, online conference, Listservs, and Newsgroups, participants can exchange their opinions and perspectives among themselves spontaneously or under the guidance of the instructors. This interactivity is believed to be a crucial function for constructing knowledge. The second category is interaction between the learners and the learning system. It is beneficial for the learners if the learning system itself is interactive, which means it is capable of providing feedback to learners promptly whenever the user input is received. This interactivity has two advantages. Firstly, in some interactive learning systems, the learners are allowed to control their learning pace, instead of passively following the uniform lecture arrangement. Individuals are assumed to be able to learn better at their learning pace since they do not have to make an effort to adjust their pace to everyone else (Zhang et al., 2006). Another advantage of interactivity is the ability to get instant feedback from the system. With this feedback, the learners can be provided with valid information about their current performance on the learning task (Aljohani et al., 2010). This feedback can also be used by the instructors as a reference to modify the learning content.

- *Authentic problem*

  One of the most critical goals of learning is to develop problem-solving abilities. Many educators believe that the problem-based learning (PBL) is the best approach to acquiring this ability. The PBL is a learner-centered approach that fosters the learner's problem-solving abilities by presenting an authentic problem and encouraging learners to solve it with an independent thinking capacity and collaborative learning. For this learning approach to work, one crucial element is the authentic problem that is as close to the realistic situation as possible. This element is usually missing in the didactic instructional learning approach, where knowledge is transferred from the instructors to the learners using abstract symbolic teaching systems. The output of a traditional didactic learning approach is abstract knowledge. There is a clear gap between abstract knowledge and problem-solving abilities. The learners with the abstract knowledge need to bridge the gap by converting what they know to what they can do to solve the practical problem. However, this process of conversion is neither effortless nor natural. Thus it is entirely possible that one may be unable to address the problem even if he/she has already mastered all the required knowledge obtained through the traditional didactic learning approach. This situation will be far less unlikely under CBL approach. With the authentic problems available during the learning process, the conversion above is no longer necessary, and the learner are able to solve the problem with both knowledge and problem-solving abilities instead of abstract knowledge alone.

- *Control*

  The degree of control provided to the learners in a learning environment can lead to increased IM and learning effectiveness. There are several kinds of control in a learning environment, including the choice of the learning path, the order of the learning activities they choose to complete, and the learning strategies to construct the knowledge. In more specific scenarios, the controllability can also include the ability to manipulate a virtual object in case of a virtual environment, or to control the behaviors of an avatar in a character-based virtual world. With the high degree of controllability, the learners feel a sense of self-determination during the learning process, and additionally be intrinsically motivated. Adversely, in a scenario where the controllability is relatively low, for example in a didactic classroom, the learners usually feel bored and reluctant to follow the predetermined lecture or the instructor's arrangement. Thus they may lose interest in the learning content quickly. Under such circumstances, extrinsic motivation, which is not as effective, is usually used to stimulate the learners to participate.

- *Challenge*

  The optimal amount of challenge can intrinsically motivate the learners to seek knowledge and explore learning content. Here this research emphasizes the optimal amount because the relationship between the amount of challenge and the IM is not as simple as being monotonic. If the challenge presented in the learning system is too low, the learners can easily become bored and quickly lose focus on the learning content. When the challenge increases gradually, the learners will start to concentrate on the learning activities again, with the urge to conquer the difficulties that come along with the challenge. With the amount of challenge rising, there is a moment when the maximum motivation and interest is inspired, and the learners feel engaged to face the challenge and to receive an ultimate sense of accomplishment once the challenge is overcome. When the challenge faced by the learners continues to rise and passes a threshold, the learners would be occupied by a strong feeling of frustration and lose all interest and motivation in further exploring the learning content. So to keep the learners in the positive zone of IM, the level of challenge is most critical.

- *Experience*

  The experience of simulative sensation is a significant source of IM that can benefit the learning effectiveness. The amount of diverse types of experiences one can engage in is always limited, especially those that are not easy to obtain in everyday life. So, if a learning system can provide the learners with the opportunity to experience something different, the learners are always motivated to participate, whether or not there is a reward attached to it. One example brought up by (Huang et al., 2013) is that to motivate a learner to study autonomy, the learning activity that allows the learners to experience freedom would most likely enhance the student's perception of self-government. For all kinds of positive emotions that play a major role in enhancing the student's learning interest, exposing them to the experiences of those positive emotions would definitely motivate the learners to participate in the learning activities, and the learning systems that are capable of providing such experiences are better choices over those that are not.

3.2.4 Research Model

A research model is a theoretical framework that proposes the relationship network among different categories of variables including independent variables, latent variables, and dependent variables. Based on the research model, mathematical model and testable research hypotheses can be generated so that the proposed theory can be validated.

According to the literature review, there are several VR factors that can impact the learning outcome. Among these candidate VR factors, four factors were selected for further research according to previously mentioned selection criteria. Moreover, the theoretical learning frameworks such as CBL and IM theory are also believed to play a role in the learning outcome. The CBL and IM can be further decomposed into critical components. All those factors have some level of impact on the final response variable - the learning outcome, directly or indirectly. To find out how these variables are affecting each other, a research model that discloses the relationship network among all levels of factors is proposed. The conceptual framework of this research model is demonstrated in Figure 3.2. Based on the conceptual framework, a multivariate regression model and a group of statistical hypotheses are proposed.

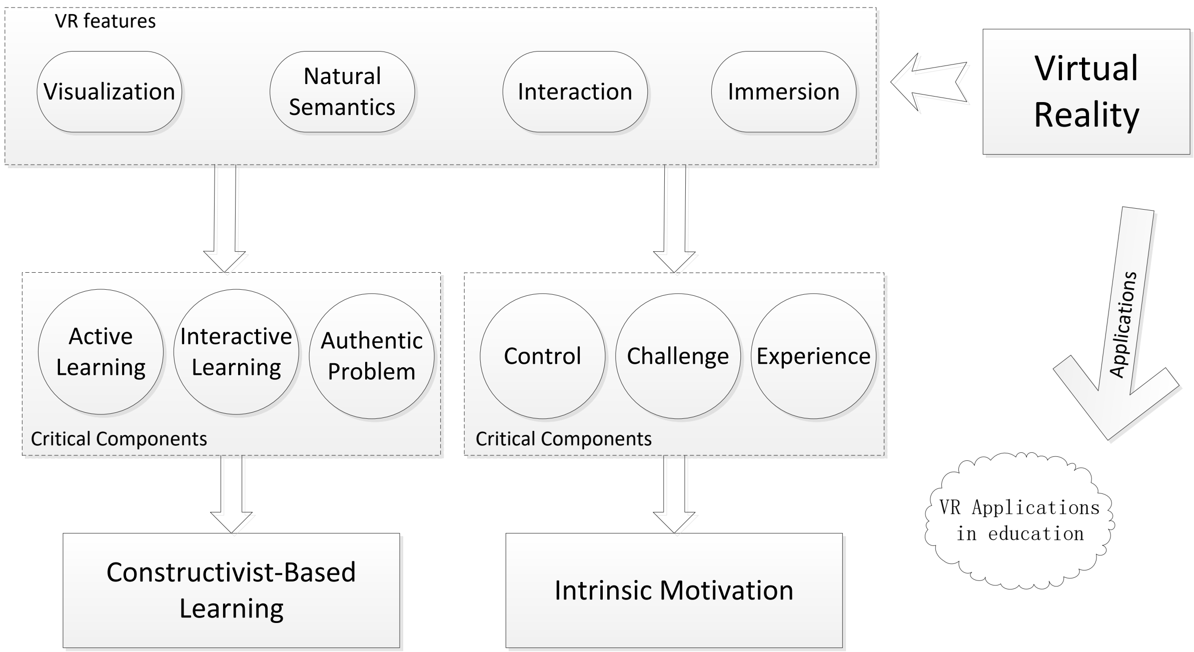

Figure 3.2: Conceptual framework of the research model

3.2.5 Multivariate Regression Model and Hypotheses

From the research model proposed in the previous section, two types of statistical instruments can be used to study the relationship between VR and the learning outcome.

Firstly, this research used a multivariate regression model to explore the quantitative relationship between each VR factor and the learning outcome. The model can be represented by Equation 3.1:

$$Y = \beta_0 + \beta_i X_i \sum \sum \beta_{ij} X_i X_j + e \tag{3.1}$$

Where Y denotes the learning outcome, X_i stands for the selected factors which are binary variables, β_i and β_{ij} stands for the coefficient of each factor and their 2-strength interactions. e is the uncounted variation. If the specific variables are plugged into X_i, the model is like.

$$
\begin{aligned}
LearningOutcome = \beta_0 + \beta_i * Visualization + \beta_2 * NaturalSemantics + \\
\beta_3 * Interaction + \beta_4 * Immersion + \\
\beta_{12} * Visualization * NaturalSemantics + \\
\beta_{13} * Visualization * NaturalSemantics + ... + \\
\beta_{34} * Interaction * Immersion
\end{aligned}
\tag{3.2}
$$

Another approach to analyzing how the VR factors can impact the learning outcome is through the critical components of theoretical learning frameworks. A list of hypotheses can be generated from the research model, and by testing these hypotheses, conclusions can be drawn on whether or not there are any significant correlations between the VR factors and the critical components. Since it is unclear what the potential correlation structure would be like, this research makes no premises, and consider all possible combinations of correlations. There are four VR factors and six critical components, thus a total number of 4*6 = 24 pairs of correlation combinations are generated, each represented by a hypothesis. The possible hypotheses are listed in Appendix A.

And finally, the grand hypothesis is made to test whether there are any significant differences in any of these group mean comparisons. The grand null hypothesis H_0 and the alternative hypothesis can be defined as below:

- *Hypothesis 1 (H_1): At least one hypothesis from H_{0a} to H_{0j} will be rejected.*
- *Hypothesis 0 (H_0): Hypotheses from H_{0a} to H_{0j} will all fail to be rejected.*

The research would fail to reject this grand null hypothesis only if all the hypotheses fail to be rejected. In other words, if at least one of those null hypotheses can be rejected, the H_0 can be rejected. Those hypotheses will be tested by comparing the group means of the critical components rating with the VR factors of interest as the grouping variables.

3.3 Development of WIPLS

The experiment is based on Walk in Place Learning System (WIPLS) with pedestrian road safety as the learning subject. The WIPLS is highly customizable where each factor being researched can be tuned at the low level or high level, thus, yielding all possible combinations of the sub VR-systems.

The WIPLS consists of the hardware component and the software component. The hardware component includes a Microsoft Kinect Sensor, a commercial TV screen, a PC, and an iPhone. The Kinect is a line of motion sensing input device developed by Microsoft for Xbox and Windows PCs. The software creates a 3-D virtual scenario based on a real suburban community and was developed using the Unity3D game engine.

3.3.1 Hardware

In this research, the Kinect is used to track the skeletal joints of a human standing in front of the sensor. 20 key joints can be detected and tracked by the Kinect (as shown in Figure 3.3). Tracking these joints renders possible the detection of various human body movements such as walking behaviors. With a capture rate of 30 frames per second, the trajectory of each joint is smoothly tracked in real time. In the Kinect system, tracking is performed by coupling RGB and depth sensors (Schalkoff, 1989). Because this research has adopted non-immersive VR technology (i.e., a screen instead of a head mounted display (HMD)), a commercial level TV is chosen as the screen to provide the virtual display. As mentioned above, the joint skeletal data collected by the Kinect sensor is used in this research instead of the raw image stream. This significantly reduces the computational load. The relatively inexpensive combination of commercial devices is sufficiently powerful to handle computational complexity while producing smooth visual feedback.

Because the subject must remain in the sensor's field of view (FOV), the WIPLS requires a human to stay in a bounded physical space. To satisfy this requirement, feedback (the display sensor's FOV at the corner of the TV screen as shown in Figure 3.4) and feedforward (the placement of a cross mark sign on the floor) were used to prevent users from leaving the sensor's FOV.

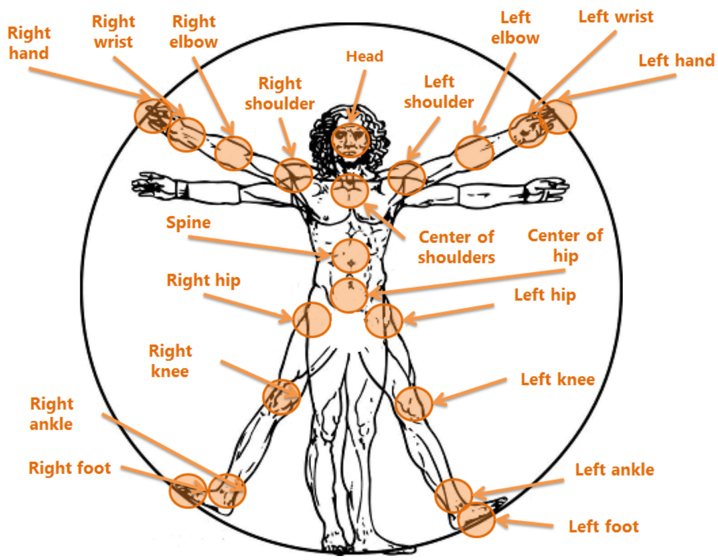

Figure 3.3: Vitruvian man and the 20 joints tracked by the Kinect

Figure 3.4: Feedback and feedforward system in WIPLS

With the Microsoft Kinect camera as the natural way of controlling method, accordingly, an iPhone 6s is used as a wireless controller for the less natural way. As displayed in Figure 3.5, an application called Joypad Legacy is installed on the iPhone, and the users can now control the WIPLS system just like a traditional video game.

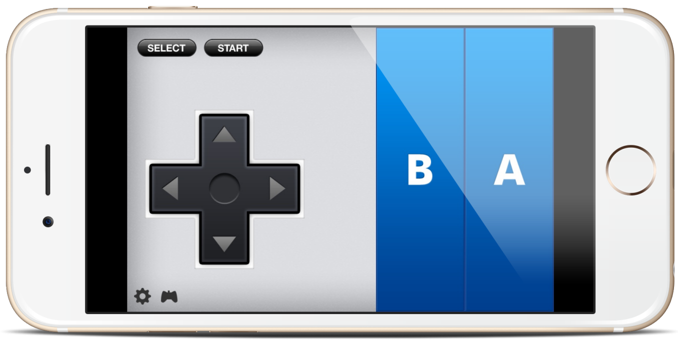

Figure 3.5: Using iPhone as a wireless controller

3.3.2 Software

Once tracking data are acquired through the hardware system, data were processed by the software system to generate smooth locomotion. Three components of the software system were discussed in this section: the zero crossing-based algorithms implementation, the speed-dampening algorithm, and rotation detection.

Zero Crossing-Based Algorithm

The joint trajectories tracked by the Kinect sensor are susceptible to variations caused by system and random errors. In this study, the zero crossing algorithm is applied to reduce this variation and accurately detect WIP steps. The zero crossing algorithm is commonly used in electronics, mathematics, and sound and image processing. This algorithm is also used in pedestrian dead reckoning, (Beauregard, 2006; Chen et al., 2010) step length estimation, (Shin et al., 2007) and step detection (Alzantot and Youssef, 2012) in pedestrian tracking technologies. The zero crossing algorithm describes a point where the sign of a mathematical function changes. It is based on the zero crossing rates (Chen, 1988) (ZCR), at which the signal changes from positive to negative or vice versa. ZCR is defined as:

$$\mathrm{ZCR} = \frac{1}{T-1}\sum_{t=1}^{T-1} \mathbb{I}\{s_t s_{t-1} < 0\} \tag{3.3}$$

$$s_t = KneeDiff = LeftKnee_t.Y - RightKnee_t.Y \tag{3.4}$$

where $\mathbb{I}\{A\}$ is an indicator function, if the argument A is true, $\mathbb{I}\{A\}$ returns 1; otherwise, it returns 0. In this study, if $s_t s_{t-1} < 0$, then $\mathbb{I}\{s_t s_{t-1} < 0\} = 1$, otherwise, $\mathbb{I}\{s_t s_{t-1} > 0\} = 0$. s_t is the knee difference at time t and s_{t-1} is the knee difference at time $(t-1)$ ($s_t = KneeDiff$ and $s_{t-1} = preKneeDiff$). When a human is walking, he/she will move by lifting and setting down each leg alternatively. This locomotion will cause s_t to change sign for each step.

Speed-dampening Algorithm

Whenever a step is detected by the Kinect sensor, a change in speed will be generated in the virtual world. In practice, there are two commonly used methods for determining forward speed (Feasel et al., 2008; Istance et al., 2009). One method is to use body position as an input and to produce keystroke and mouse events as outputs (Istance et al., 2009). For example, when the subject presses and holds the "forward" arrow key on the keyboard, he/she in the virtual world will keep moving forward until he/she releases the "forward" arrow key. The advantage of this method is that it is simple and straightforward and does not require changing the system configuration. With this approach, the stepping event is treated as a hardware interrupt event. The disadvantage of this method is that the frequency of the step event (about 2Hz) is much slower than the frequency of hardware interrupt events (about 100 Hz). As a result, there are few speed impulses in each second, which will certainly lead to severe jerkiness during walking. An alternative method is to use the box and the saw-tooth functions as applied in the LLCM-WIP system (Feasel et al., 2008). Using this approach, the jerkiness between the two consecutive impulses can be smoothed. This study uses a revised saw-tooth function for speed smoothing. In each frame, the function

SmoothDamp() is called to dampen the speed from the current value to 0 within a short period of time (e.g., 0.5 seconds). If the user stops generating new speed increments, the advancement of the viewpoint in the virtual world will stop after 0.5 seconds. If the user is continuously walking, the acceleration from the ZCB algorithm will counteract the deceleration from the speed-dampening algorithm, such that the speed of the subject in the virtual world is relatively stable and continuous. To summarize the speed-dampening algorithm, KneeSwap increases while a step is detected and reaches 0 in 0.5 seconds if there is no step detected. The 0.5 seconds is also selected empirically.

Additionally, to avoid abrupt speed changes, a 4-period moving average speed is used to smooth the most recent speed values and reduce unwanted randomness and period-to-period speed variations. The speed changes before and after smoothing are described in Figure 3.6. This research concludes that whenever the knee difference results in a zig-zag pattern (green dotted curve) indicating that the subject is walking, the raw speed will gain an increment (blue dashed curve). It is also worth mentioning that because of the nature of the ZCB algorithm, the magnitude of the knee difference has no direct impact on the walking speed. By applying the 4-period moving average, the variation of the smoothing speed becomes small (red solid curve). This smoothing speed will finally drive the advancement of the viewpoint and enable the subject to move in the virtual world.

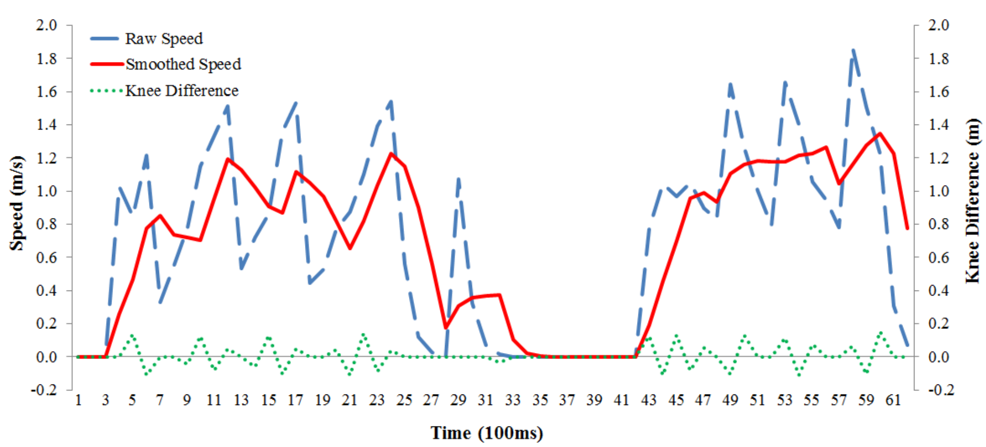

Figure 3.6: Knee difference and locomotion speed

Rotation Detection

The joint position data of the left shoulder and right shoulder collected by the Kinect are used to track the subject's rotation. As discussed above, the Kinect sensor can also capture the depth value of each pixel as well as each body joint. As seen in Figure 3.7, when a human turns left or right, the depth value of the left shoulder and right shoulder joints will increase and decrease, respectively. When turning left, the difference between the depth values of the left and right shoulder joints ($LeftShoulder.Z - RightShoulder.Z$) will change from 0 to a positive value. Similarly, if the subject turns to the right, this depth difference will change from 0 to a negative value. In order to tell the real turning behavior, another threshold value is set. If the absolute value of the depth difference is smaller than the threshold value, it is safe to consider this depth change as noise. The turning angle is directly proportional to the depth difference, meaning that the more the subject is turning apart from looking straight, the greater viewing angle change will be displayed in the VR system. When the subject is facing the screen directly, there will be no displacement of viewpoint in the VR system.

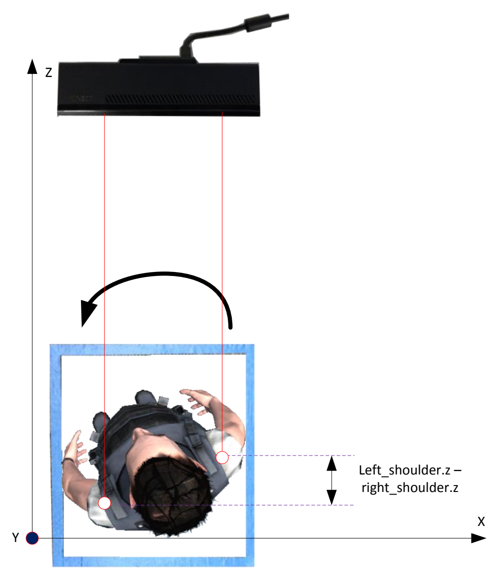

Figure 3.7: Bird eye view of making a left turn

3.3.3 Pilot Studies for Apparatus Evaluation

To evaluate the performance and features of the WIPLS, an objective experiment and a subjective survey are conducted.

Experiment One: Objective Performance Evaluation

Seventeen participants (11 male; 6 female) were recruited to participate in the evaluation experiment. These participants were selected from a convenience sample of students and staff at the University of Tennessee. Their ages range from 21 to 38, and their height varies from 65 to 74 inches. The majority of participants had little or no prior VR gaming experience before using the WIPLS, and some had only limited knowledge of Microsoft Kinect gaming. Prior to using the WIPLS, participants were not informed of the purpose of the experiment.

This research used a simple evaluation program based on the ZCB-WIP system to evaluate the actual latency from the participants' performances. The study participants were asked to follow the instructions on the screen, such as "GO" and "STOP" with a downward counting timer (see Figure 3.8). During the experiment session, three variables are recorded: (1) the value of knee difference captured by the system; (2) the immediate locomotion speed before smoothing; and (3) the locomotion speed after smoothing. The sampling rate for these variables is 10Hz; thus, each data point represents 100ms. This research chose a moderate sampling rate instead of a higher value mainly for performance considerations. Variables are stored in a local file for post-processing and statistical analysis; thus, increasing the sampling rate results in I/O operations that may bring extra load on the computer and adversely impact the framerate of the visual feedback. Also, according to the result of the analysis, 100ms is an acceptable level of granularity for the study. The latencies can be calculated by counting the number of data points. Both the starting latency and the stopping latency are calculated.

Experiment two: Subjective Survey Analysis

In addition to the objective experiments, a second group of participants was recruited for a subjective system evaluation. This group included 35 participants (29 male; 6 female), aged 13

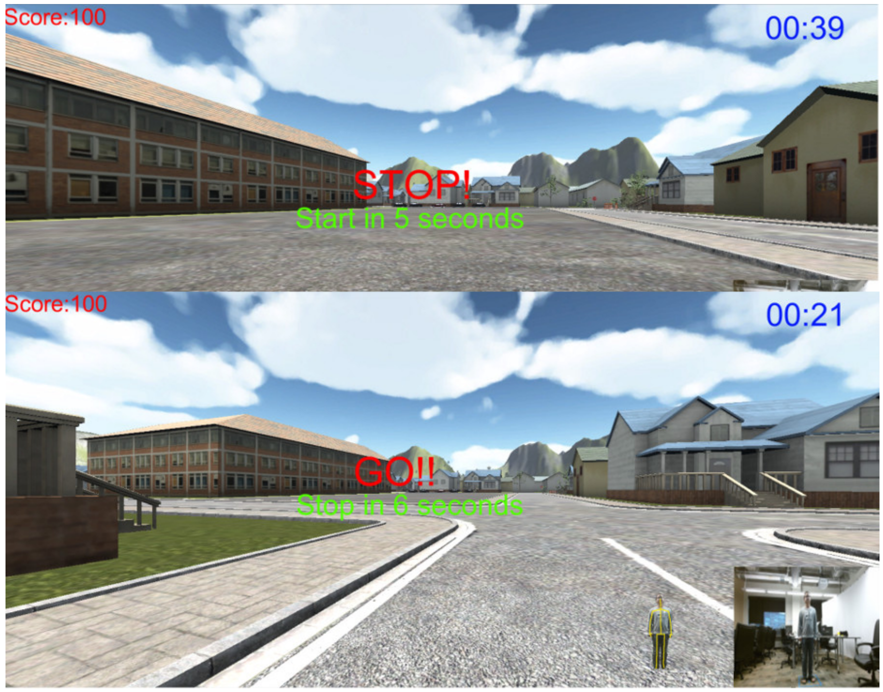

Figure 3.8: Instruction text for users to start or stop walking in place

to 17 years. The group of participants was asked to experience two VR systems; one is the WIPLS and the other is a demo program using the Oculus Rift HMD and traditional keyboard/mouse control. Eight subjective survey questions were answered by the participants after they tried both VR systems to rate their subjective experiences while using each VR system. These eight survey questions were selected from well-known VR evaluation questionnaires (Witmer and Singer, 1998), with proper modification and rewording. The specific question items are listed below.

- Q1: Walking is natural or not? Scale: 1 is most artificial and 5 is most natural.
- Q2: System is responsive or not? Scale: 1 is not responsive and 5 is most responsive.
- Q3: How much fatigue do you feel during the experiment session? Scale: 1 is least fatigue and 5 is most fatigue.
- Q4: How much motion sickness do you feel during the experiment session? Scale: 1 is least motion sickness and 5 is most motion sickness.
- Q5: How much latency (lag) do you feel during the experiment session? Scale: 1 is least latency and 5 is most latency.

- Q6: How much immersion (being there) do you feel? Scale: 1 is least immersive and 5 is most immersive.
- Q7: How much easiness is the virtual system to you? Scale: 1 is very easy and 5 is most complicated.
- Q8: How much comfort do you feel when experiencing the system? Scale: 1 is not comfortable and 5 is most comfortable.

3.4 Design of Experiment

In this section, the Design of Experiment (DOE) is introduced as a statistical tool to design the experiment combinations and implement sub-VR systems for this research.

3.4.1 Introduction to DOE

DOE is a systematic method to determine the relationship between factors affecting a process and the output of that process (Anderson and Whitcomb, 2016). It functions by manipulating the levels of one or more controllable input factors and observing the corresponding response variables in order to find the cause-and-effect relationship.

The most straightforward way of designing an experiment is using full factorial design because it is easy to design, efficient to run, and contains abundant information to support plenty of statistical analysis like Analysis of Variance (ANOVA) or factorial analysis. Despite the significant number of advantages, the biggest drawback is that the full factorial designs always require a huge number of treatment combinations as well as experimental runs, thus resulting in a rather high experiment cost. This problem is even more severe when involving human subjects since it may either prolong the experimental session in a within-subject design, or require a considerable number of participants in a between-subject design. The former issue may increase boredom to the participants and affect the accuracy of the results, and the latter may jeopardize the statistical power of research when the number of recruited participants is not sufficient.

A good alternative to avoid the dilemma is to use fractional factorial design instead. Compared to a full factorial design, a fractional factorial design permits the investigation of the effects of many factors in fewer runs. The reduced number of treatment combinations in fractional factorial design will bring confounding structures between the main effects and some interactions, but this cost is usually acceptable because any interactions involving three factors and higher order are unlikely to impact the response variable significantly.

The objective of the research design is to find out how each factor can impact the learning effectiveness, both individually and through interaction. Since there are four factors with two levels in each factor, for a full factorial design, a total number of $2^4 = 16$ combination treatments is required, which will be both unnecessary and cost inefficient. The ability to measure three-way interaction and four-way interaction does not provide enough meaningful insights to this study, and the statistical power will be greatly compromised. Under such circumstance, the fractional factorial design is an excellent choice.

3.4.2 Application of DOE in Research

To determine how many combination treatments are suitable for this fractional factorial design; this research conducted the DOE using JMP's Custom Designer. As displayed in Figure 3.9, four categorical variables from X1 to X4 were chosen; each variable contains two levels, L1 and L2. The Custom Designer then recommends using 12 runs for this experiment.

After clicking the 'Make Design' button, the actual combination treatments can be generated, which is listed in Table 3.3. This design is able to estimate the main effects and second-order interactions between all these main effects. From Figure 3.10, it is easy to tell that this design has a high D Efficiency, G Efficiency and A Efficiency, indicating that the goodness of the design is maximized. With the experimental design listed in Table 3.3, it is able to plug in the real factors and the corresponding levels into the design. Table 3.4 describes how the low level and high level (L1 and L2) of each factor is designed and manipulated in the customizable VR system.

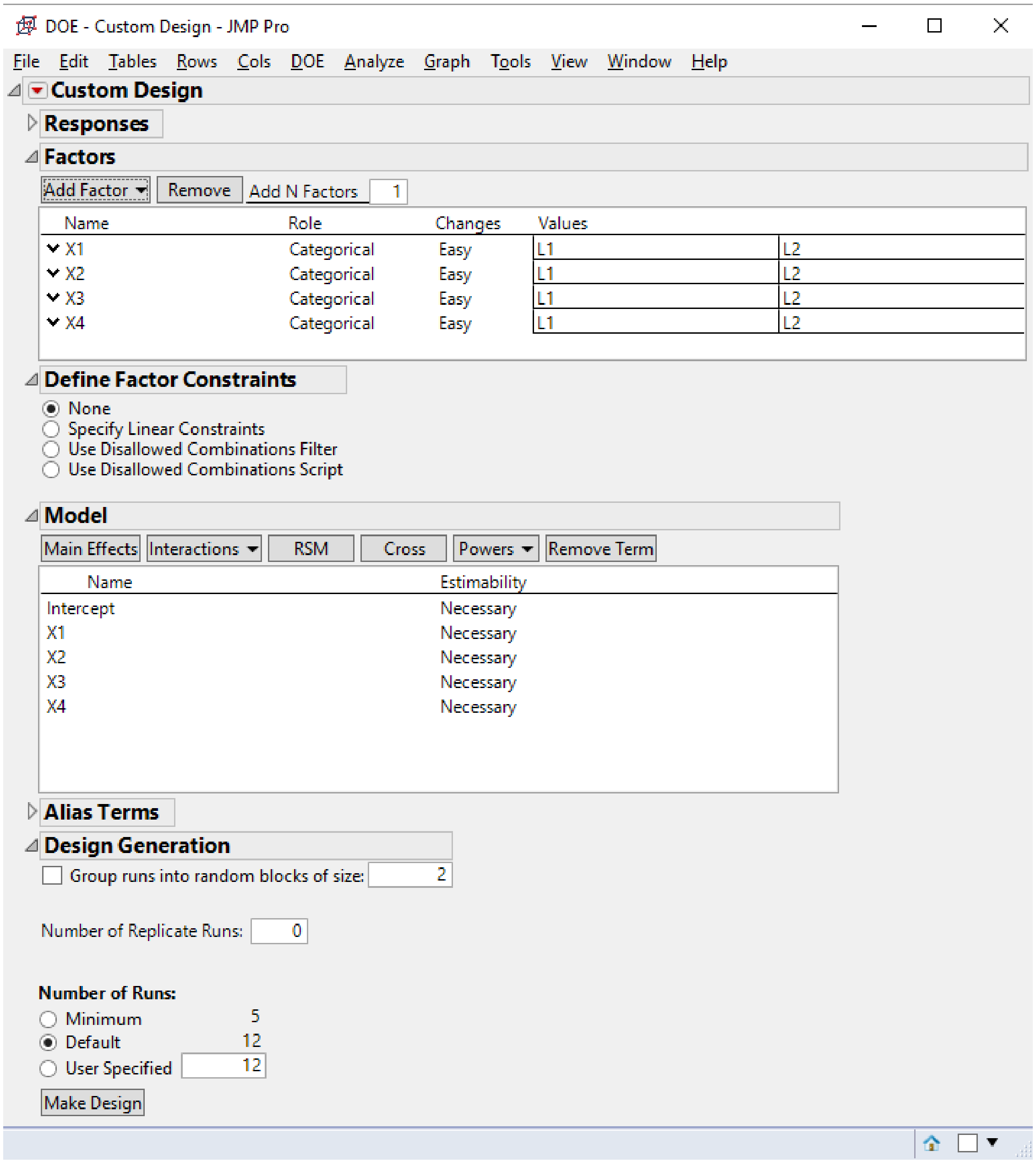

Figure 3.9: Design of Experiment: Custom designer in JMP

Table 3.3: Treatment combinations with 12 runs and 4 two-level factors

Run	X1	X2	X3	X4
1	L2	L2	L2	L2
2	L1	L2	L2	L2
3	L1	L1	L1	L1
4	L1	L1	L2	L1
5	L2	L2	L1	L1
6	L2	L1	L1	L2
7	L2	L2	L1	L1
8	L2	L1	L2	L2
9	L1	L2	L1	L2
10	L1	L1	L1	L2
11	L1	L2	L2	L1
12	L2	L1	L2	L1

Design Diagnostics

D Optimal Design

D Efficiency	100
G Efficiency	100
A Efficiency	100
Average Variance of Prediction	0.416667
Design Creation Time (seconds)	0

Figure 3.10: Design diagnostics of the custom design

Table 3.4: Interpretation of low level and high level for each factor

Label	Factor	L1	L2
X1	Visualization	Low level of visualization will set the display in black and white	High level of visualization will use full color in display
X2	Natural Semantics	With low level of natural semantics, the user will use a traditional game controller as the control system	With high level of natural semantics, the user will use his body language as the control system
X3	Interaction	With low level of interaction, the user will have no ability to interact with the VR system except pausing/resuming the automatic play	With high level of interaction, the user will have full control over the VR system, making any decision at any time
X4	Immersion	Low level of immersion will display the VR system using a low Field of View (FOV = 60)	High level of immersion will show the VR system using a high Field of View (FOV = 90)

The specific design table after the factors have been plugged in is listed in Table 3.5. Each treatment combination is corresponding to a sub-VR system. The characteristic of each sub-VR system is also introduced in the last column of this table.

Table 3.5: Specific design with treatment combinations and sub-VR systems

Sub-VR System	X1 Visualization	X2 Natural Semantics	X3 Interaction	X4 Immersion	System Description
1	L2	L2	L2	L2	Colored, body control, High FOV, game
2	L1	L2	L2	L2	Black/White, body control, High FOV, game
3	L1	L1	L1	L1	Black/white, hand controller, low FOV, video
4	L1	L1	L2	L1	black/white, hand controller, low FOV, game
5	L2	L2	L1	L1	Colored, body control, low FOV, video
6	L2	L1	L1	L2	Colored, hand controller, High FOV, video
7	L2	L2	L1	L1	Colored, body control, low FOV, video
8	L2	L1	L2	L2	Colored, hand controller, HIGH FOV, game
9	L1	L2	L1	L2	Black/White, body control, High FOV, video
10	L1	L1	L1	L2	Black/White, hand control, High FOV, video
11	L1	L2	L2	L1	Black/White, body control, Low FOV, game
12	L2	L1	L2	L1	Colored, hand control, Low FOV, game

3.5 Survey Instrument

The survey instrument was derived from previous literature on different areas of research including learning effectiveness, virtual reality, and learning frameworks (constructivist-based learning and intrinsic motivation). The survey instrument covered the participants' demographic information and background information regarding video games and VR. The main part of the survey questions is 18 Likert-scale questions regarding the participants' perceived learning effectiveness, satisfaction, and critical components of the learning frameworks. Those Likert-scale questions measure the participants' perceptions of learning outcome with 7-point scales, ranging from 'strongly disagree' to 'strongly agree'. The survey instrument is presented in Appendix B.

3.5.1 Survey Instrument

A pilot study was administered in June 2016. The purpose of this pilot test was to test content validities of the survey instrument and the experiment procedure. Below is the process and results of the pilot study.

Participants

The participants involved in this pilot study were 28 high school students from Knoxville, Tennessee. They are participating in a Kids U Summer Camps program hosted at the University of Tennessee, Knoxville. This program includes a broad range of activities, and this pilot study is only one of them. Among those participants, 82.1% are male and 17.9% are female. All students signed up to participate in the project voluntarily.

Data Collection

The data collection process lasted for four days within the same week. For each day, a group consisting of 6-8 students was invited to the Natural Interaction Lab in the Department of Industrial and System Engineering to participate in the project. In the beginning, the researcher introduced how this experiment would be arranged; then each participant would perform the default sub-VR system that had all features enabled. The purpose of this practice session was to let the participants have a general idea of how the WIPLS works. Following the practice session was the actual treatment session. In this session, each participant was assigned to a particular sub-VR system from the list of all sub-VR systems in Table 3.5. After the two sessions were finished, the participants finally completed a survey regarding their opinions about the treatment session and the learning effectiveness they perceived from the Virtual Reality Learning System. The survey is presented in Appendix B. Demographic questions like age and gender were asked in this survey, but no identifiable information was requested. The participants were allowed to decline or end the participation at any time during the experiment, and they could ask the researcher any questions during the whole sessions.

3.5.2 Measurement

The survey instrument used in this research is validated regarding reliability and validity. Reliability is the overall consistency of a measure. A high reliability value means that repetition under the same condition will always produce similar results. Reliability is usually measured using Cronbach's alpha (Cronbach, 1957), which is a coefficient to measure how closely a set of items are as a group. This pilot study uses this coefficient to test the scale reliability of the items in the survey instrument. The Cronbach's alpha is calculated as below:

$$\alpha = \frac{N\bar{c}}{\bar{v} + (N-1)\bar{c}} \tag{3.5}$$

Where N is the number of items, $\bar{c}$ is the average inter-item covariance among the items and $\bar{v}$ is the average variance.

The α coefficient of reliability ranges from 0 to 1.00, providing the assessment of internal consistency among all items comparing to the overall scale. If all the scale items are entirely independent of one another, then $\alpha = 0$; high covariance among all items will yield a large α coefficient, and this coefficient will increase along the number of items N, approaching 1.00 with N approaching infinity. In conclusion, the higher the coefficient is usually indicates a highly reliable instrument. There is different literature about the acceptable values of α, most where the researchers consider a coefficient that is higher than 0.70 to be acceptable.

3.6 Empirical Case Study

In this section, the empirical case study was introduced. The empirical study was used to evaluate the conceptual framework and answer the research questions.

3.6.1 Participants

The sample population is college students from both undergraduate and graduate student categories. Convenience sampling was selected as the sampling method because the students were attending the Lean Enterprise Systems Summer Program (LESSP) in July 2016 at the University of Tennessee, Knoxville at the Department of Industrial and System Engineering. The participants were accessible to the researcher and qualify as target subjects of the research. The participants were invited to participate in the experiment process in the Natural Interaction Lab in the Department of Industrial and System Engineering during their presentation week in the middle of July 2016. All participants were informed that their participation was voluntary and they could contact the researcher or the University of Tennessee, Knoxville Institutional Review Board (IRB) Compliance Officer for any questions related to this study.

3.6.2 Ethical Considerations

Consent forms were required for this research since the participants were required to finish the survey after they completed the experimental process, and the survey gathered information from the students. Consent was essential for participating in the survey. The informed consent form was provided so that no identifying information was collected with the data. The participants' data was collected anonymously so that the possible risk of a confidentiality leak was unlikely to occur, and their participation would have no impact on their academic performance in the LESSP. A hard copy of the informed consent form was signed by each participant and they were offered the opportunity of taking a signed copy of their informed consent form. Data collection procedures for the survey deployment were approved by IRB, University of Tennessee, Knoxville (See Appendix C).

3.6.3 Experiment and Data Collection Procedure

The experiment procedure is similar to the pilot study. All participants took three sessions, which were: practice session, experiment session, and survey session. The details of those three sessions are explained in the following sections.

Practice Session

The main difference from the pilot study lied in the practice session. The pilot study used one of the treatment VR systems (the treatment with all factors in high level) for the students to get familiar with the WIPLS. While this approach of practice was straightforward, there were several drawbacks. Firstly, since the practice session was too similar to the forthcoming experiment session, it caused some confusion to some of the participants, as they were not sure which session they were going to evaluate in the survey. Secondly, for some participants, the treatment VR system assigned to them in the experiment session was identical with the practice session, which means those participants would have to perform the same treatment VR system twice, causing some unnecessary bias. Lastly, the treatment VR system used in the practice session contained only high levels of each factor. For example, the factor of natural semantics was implemented with Kinect controller as the high level and the traditional gaming controller as the low level; so everyone used Kinect as the controller while nobody used a traditional game controller. This lost the point of letting all participants get familiar with the WIPLS because the researcher had to teach some of the participants how to use traditional gaming controller to control the game during the experiment session.

To properly tackle these issues, this research designed a brand new practice session in the experiment procedure of the sample population. As demonstrated in Figure 3.11, the new practice session used a straight road of a different street block in the same virtual world with the treatment VR systems. The visual theme was similar to those used in the experiment session but caused no confusion to the subjects. The subjects' objective in the practice session was to follow the audio guidance from the program and walk to the end of the road. The audio guidance asked the subject to walk forward and make turns by using walking in place behavior or using the controller on the iPhone. The corresponding text was also displayed at the top of the screen. An intuitive indicating image on the up right corner gave visual hints to the subjects to make sure that they understood what to do.

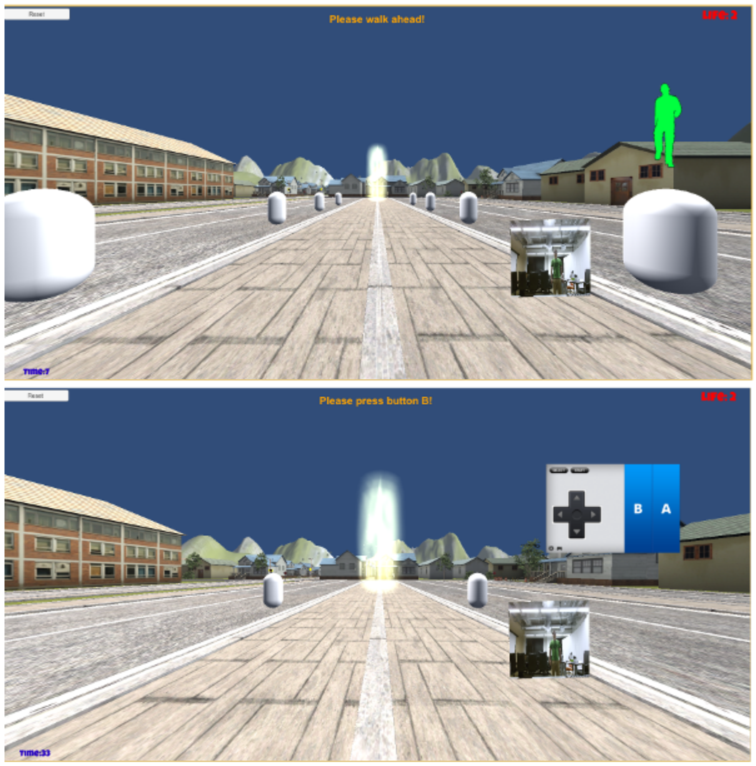

Figure 3.11: Practice session in experiment procedure

Experiment Session

The experiment session of this research used pedestrian road safety as the learning subject. The participants would learn how to cross the road safely using the WIPLS. There were four different scenarios of crossing the road in a virtual world of WIPLS, emphasizing different skill sets of road crossing. These four different scenarios are listed as follows:

- *Scenario 1:* An empty street with no traffic coming through will be presented. The participant will follow the voice instruction to look left, right, and look left again before crossing the road and arriving at the other side.
- *Scenario 2:* A street with two-way traffic and a zebra zone will be presented. The difference between this scenario and Scenario 1 is that a zebra zone is used to provide some protection to the participant; when the participant enters the zebra zone while no vehicle is inside the zebra zone already, the vehicle will stop and wait for the participant until the participant walks outside of the zebra zone.
- *Scenario 3:* A street with two-way traffic and no zebra will be presented. The difference between this scenario and Scenario 2 is that there is no zebra zone in the street, so the participant needs to look out and avoid being hit by any moving traffic.
- *Scenario 4:* A street with two-way traffic and parked cars on the roadside. The difference between this scenario and Scenario 3 is that there is a line of vehicles parked along one side of the road which limits the visibility of the participant. Before the participant starts to look left, right, and left again to check for moving vehicles, he/she has to step ahead to the edge of the parked vehicles, then look out as well as avoid the traffic on the street.

The demonstration of these four scenarios can be found in Figure 3.12. Those four scenarios were presented to each of the participants in the experiment session. For the participants with treatments containing a high level of the factor interaction, they need to avoid being hit by the moving vehicles in all scenarios except Scenario 1, in which there was no traffic. For the participants with treatments containing a low level of the factor interaction, since they were controlling a pre-recorded video instead of the VR game, they had no control over the actions of the crossing behavior, and there was no chance of being hit by vehicles.

Survey Session

Right after a participant finishes his experiment session, he/she was asked to complete a survey from either a PC or a laptop in the same lab. The questionnaire can be found in Appendix B. Demographic information including age and gender were asked besides the experiment related

Figure 3.12: Four scenarios of the pedestrian road safety learning environment

questions. The VR treatment number was also asked at the very beginning to indicate which VR treatment the current participant belonged to. An instructor was always available to answer any questions raised by the participants through the whole survey session.

3.7 Statistical Methods

In this section, the statistical tools for the data analysis are introduced.

3.7.1 Exploratory Factor Analysis

This research planned to measure the learning outcome using the VR system as the learning tool. The learning outcome is interpreted as perceived learning effectiveness and satisfaction. To measure these two metrics, several question items were designed in the survey, with 8 items measuring the perceived learning effectiveness and 4 items measuring the satisfaction. Although those items are extracted from previously published literature, it is still necessary to explore the underlying structure of the model and the relationship between the survey items and the measured factors in this research. To achieve this objective, Exploratory Factor Analysis (EFA) is performed.

EFA is a statistical approach for identifying a structure that underlines the relationship among a set of observed variables. It looks for variables that not only correlate highly within a group of other variables, but also correlate poorly with variables outside of that group (Field, 2009). With this technique, this research can transform the correlations among a set of observed variables into a smaller number of underlying factors, which contains all the essential information about the linear interrelationships among the original test scores. EFA has several applications, like: exploring a data set to reveal certain patterns when the researcher is unclear about the structure of the data and reducing a significant number of variables into a smaller number of factors that are more manageable. EFA can also be used to test whether a set of items designed to measure certain variables do reveal the hypothesized factor structure. This research uses EFA to explore how the survey items are correlated with the measured factors regarding learning outcome, and the factor loadings can also be used to impute the factor composites. Such composites can serve as dependent variables when exploring the quantitative model involving the VR factors and learning outcome.

In this research, the principal components method with Promax rotation was used in EFA to assess if the measured factors are in line with the survey items. Principal component is a factor extraction method used to form uncorrelated linear combinations of the observed variables. This method is variance-based; the first component has maximum variance, and the successive components explain smaller portions of variance progressively. In order to determine the appropriateness of proceeding with EFA, the Kaiser-Meyer-Olkin (KMO) measure and Bartlett's test of sphericity were performed. KMO is used to represent the ratio of the squared correlation between variables to the squared partial correlation between variables. The KMO is a ratio that varies between 0 and 1. A value of 0 indicates diffusion in the pattern of correlation, while a value close to 1 indicates that pattern of correlations are relatively compact, so factor analysis should yield distinct and reliable factors. It is recommended that a value that is greater than 0.5 can be barely acceptable, and values above 0.90 are considered excellent. Bartlett's test of sphericity examines whether the correlation matrix is proportional to an identity matrix where the values on off-diagonals are zero, and the values along diagonals are roughly equal (Field, 2009). The

Bartlett's test has to be significant in order to run a meaningful EFA. In this research, the following criteria were applied to remove survey items when identifying the factor structure from a pattern matrix in an EFA:

- Items that do not load on any factors will be removed.
- Items with low loading (less than 0.5) will be removed.
- When there are cross-loadings between factors, i.e. one item is loading on more than one factor, the primary loading should be at least 0.2 larger than the second loading. Otherwise, the item should be removed.
- An item will be removed if it loads on a factor where it is theoretically unreasonable for that item to be associated with other items in the same group.

Confirmatory Factor Analysis (CFA) is the next step of Exploratory Factor Analysis (EFA) to determine the factor structure. After the factor structure was explored in EFA, the next step is to confirm the factor structure extracted from EFA. SPSS AMOS is used for this purpose.

The purpose of CFA in this study is to collapse the survey items in each group and impute the composite factors. After importing the pattern matrix and the sample data into AMOS, it is now possible to calculate the estimates of the regression weights and impute the composite factors using the regression imputation. The imputed composite factors can then be used as the dependent variables for further analysis and investigation.

3.7.2 Aligned Rank Transfer for Factorial Analysis

In human-computer interaction (HCI) research, nonparametric data response data are frequently generated, like error counts, Likert-scale response, preference tallies, etc. For those types of data, normality is not assumed. Since normality is one of the assumptions to use ANOVA, common ANOVA cannot be applied to those types of data directly. The response variable in this study is the feedback data in Likert-scale format, which is categorical data and common parametric ANOVA can not be applied directly. Thus, we resort to statistical methods that are suitable for these kinds of data types. There are several statistical methods for handling nonparametric data, as listed in Table 3.6 (Wobbrock et al., 2011).

Table 3.6: Some possible analyses for nonparametric data (Wobbrock et al., 2011)

Method	**Limitation**
General Linear Model (GLM)	Can perform factorial parametric analyses, but cannot perform nonparametric analyses.
Mann-Whitney U, Kruskal-Wallis	Can perform nonparametric analyses, but cannot handle repeated measures or analyze multiple factors or interactions.
Wilcoxon, Friedman	Can perform nonparametric analyses and handle repeated measures, but cannot analyze multiple factors or interactions
χ^2, Logistics Regression, Generalized Linear Models (GZLM)	Can perform factorial nonparametric analyses, but cannot handle repeated measures.
Generalized Linear Mixed Models (GLMM), Generalized Estimating Equations (GEE)	Can perform factorial nonparametric analyses and handle repeated measures, but are not widely available and are complex.
Kaptein et al.'s nonparametric method	Can perform factorial nonparametric analyses and handle repeated measures, but requires different mathematics and software modules for each type of experiment design.
Aligned Rank Transform (ART)	Can perform factorial nonparametric analyses and handle repeated measures. Requires only an ANOVA after data alignment and ranking, provided for ARTool or ARTweb.

This research measures four factors and six two degree interactions with nonparametric data as the response variable, where most of those listed methods are not suitable. While the Aligned Rank Transform (ART) satisfies all the requirements in this research and is also highly accessible, this procedure is used for the factorial analyses.

The ART procedure does not convert the nonparametric data into parametric data directly. Instead, the procedure works in two main processes: the "align" process and the "rank" process. The "align" process applies some calculation algorithms on the response variable with regards to each term. Here the "term" refers to main effects and the interaction effects. After the alignment, this procedure obtains a new column for each term, in total $2^N - 1$ columns, where N is the number of factors in the model. The "rank" process then sorts each column and assign the ranks to a new column, with averages in case of ties. This produces additional $2^N - 1$ columns. These columns are used as the parametric response variables when fitting the conventional ANOVA model. Note that there are now $2^N - 1$ response variables, with each response variable corresponding to only one term, so accordingly there are $2^N - 1$ ANOVA models; while in each of these models, only the results (sum of square, degree of freedom, F value, p-value, etc) with the corresponding term is examined, the rest of the results are all ignored.

The detailed procedure is explained in more details in five steps:

- *Step 1:* Compute residuals.
- *Step 1:* Compute estimated effects for all terms, i.e. main effects and interaction effects.

  For main effects:

$$\text{estimated effect} = \overline{A_t} - \mu \tag{3.6}$$

  For two-way interactions:

$$\text{estimated effect} = \overline{A_i B_j} - \overline{A_i} - \overline{B_j} + \mu \tag{3.7}$$

  For three-way interactions:

$$\text{estimated effect} = \overline{A_i B_j C_k} - \overline{A_i B_j} - \overline{A_i C_k} - B_j\bar{C}_k + \overline{A_i} + \overline{B_j} + \overline{C_k} \tag{3.8}$$

For N-way interactions:

$$
\begin{aligned}
\text{estimated effect} &= \overline{N-way} \\
-\sum(\overline{N-1way}) + \sum(\overline{N-2way}) - \sum(\overline{N-3way}) &+ \sum(\overline{N-4way}) \\
&\dots \\
-\sum(\overline{N-hway})//&ifhisodd, or \\
+\sum(\overline{N-hway})//&ifhiseven
\end{aligned}
\tag{3.9}
$$

- *Step 3:* Compute aligned response Y'. The aligned response value is calculated as:

$$Y' = \text{residual} + \text{estimated effect} = \text{result from step 1} + \text{result from step 2} \tag{3.10}$$

- *Step 4:* Assign averaged ranks Y''. Assign averaged ranks of each value in Y' to a new column to create Y''. The smallest Y' yields 1 in Y'', the next smallest Y' yields 2 in Y'', and so on until the largest Y' yields r in Y'', where r is the number of rows in the dataset. In case of a tie among k values of Y', the value in Y'' is the averaged value among those k ranks.
- *Step 5:* Perform a full factorial ANOVA on Y''. Now Y'' is produced by the ART, it is ready to perform the conventional ANOVA using this column as the response variable. All main effects and interaction effects should be included in the model, while only the result corresponding to the effect that yields Y'' should be considered.

For a model with N factors, the procedure from step 1 to step 5 are performed for 2^{N-1} times if performing a full factorial design. To simplify this tedious process, the ARTool is used to do the alignment and transformation automatically.

3.7.3 Group Analysis on Background Variables

To investigate difference of the learning outcome between groups divided by background variables, this research used Kruskal-Wallis H test to compare the independent groups. The background information collected in the survey is used as the group variables, including gender, gaming experience, and prior VR experience. The Kruskal-Wallis H test was chosen to analyze the data since it is the nonparametric alternative of one-way ANOVA, and also the extended alternative of the Mann-Whitney U test since it allows comparison of more than two individual groups.

3.7.4 Group Means Comparison for Hypotheses Tests

The next step of analysis of the survey data is to compare the group means to test the hypotheses, so as to answer the research questions: Do the VR factors improve the learning effectiveness through the theoretical learning frameworks? Does the WIPLS support the constructivist-based learning approach and increase the intrinsic motivation of the participants? How does each VR factor fit in the theoretical learning frameworks by correlating with the critical components?

The last set of questions is used to explore the relationship between the VR factors and the critical components of the theoretical learning frameworks. The sample population can be divided into groups of equal size using each of the VR factors as the grouping variable. For example, all participants in the treatment VR systems having high level of Visualization (treatment 1, 5, 6, 7, 8, 12) form a group with exactly the same size of the group having low level of Visualization (treatment 2, 3, 4, 9, 10, 11). This is exactly half of the sample population. Using this grouping criterion, total four pairs of groups were produced, with each pair of groups covering the total sample population.

Since the two groups in each pair are from different participants taking the experiments individually, and the response variables are Likert-scale values, the best option for the analysis is to use the nonparametric equivalent test of the unpaired t-test, which is the Mann-Whitney U test.

3.7.5 Use Information Criteria for Model Selection

Besides using ANOVA to test the significance of the VR factors, an alternative approach to solving the multivariate regression model is to use the model selection method based on information criteria. This approach is to select the model from a set of competing models that best describes the underlying process of the dataset. The selected model needs to maximize the goodness to fit, which means, it should account for most variance. While if we consider goodness to fit as the sole criterion, we would end up getting a model that is too complex and generalizes poorly. This overfitting effect can be offset by using the information criteria.

There are several criteria for model selection, like the AIC (Akaike, 1998), Cp (Mallows, 1973), and BIC (Schwarz, 1978). It is also recommended to use AICc (Anderson and Burnham, 2002) instead of AIC when the sample size is small. ICOMP (Bozdogan, 1987b, 1988, 1990, 1987a; Bearse and Bozdogan, 1998; Bozdogan and Haughton, 1998) is another criteria that is based on AIC while measuring the complexity of the model differently. In this research, the model selection using AIC, BIC, AICc, Cp and ICOMP is presented together with the ART approach, and insights are obtained by comparing these different approaches.

3.8 Summary

This chapter presented the structure of this research. To explore how Virtual Reality can affect the learning outcome, a research model involving all VR factors and the dependent variable was proposed. The latent variables representing the theoretical learning frameworks indirectly affecting the learning outcome were also included. A multivariate regression model and a list of research hypotheses were also presented to test the quantitative relationships between these variables in the research model.

Next, the experiment apparatus, the Design of Experiment, and the survey instrument were presented. The technical details of the experiment apparatus including the hardware and software implementation were discussed, including the algorithms developed for the performance optimization. A pilot study was also designed with the purpose of evaluating the objective performance and subjective feedback of the experiment apparatus. The Design of Experiment was also discussed, covering both the general introduction and the specific treatment combinations in this research. After that, the survey instrument extracted from literature review was presented, and the second pilot study was also introduced with the objective of evaluating the internal consistency and the validity.

With the sub-VR systems implemented using the WIPLS guided by the results of the DOE, an empirical case study was conducted next to evaluate the research model. The experiment procedure was discussed, including a practice session, an experiment session, and a survey session. Finally, the statistical methods which will be used in Chapter 4 for experiment data analysis were described. The specific data analysis for the pilot studies as well as the empirical case study will be discussed in detail in chapter 4.

Chapter 4

Research Findings

4.1 Overview

Two pilot studies and one empirical case study were conducted in this research. The data results of those three studies and the analysis is presented in this chapter. The data analysis method includes descriptive statistics, frequencies, correlation analysis, and factorial analysis in finding a relationship between variables. Since this research included response variables in discrete format and small sample size in pilot studies, normal parametric analyses are not applicable. Thus, nonparametric statistical methods for the data analysis are applied.

4.2 Pilot study 1: WIPLS Evaluation

As mentioned in chapter 3, two experiments were conducted in the first pilot study to evaluate the performance of the WIPLS and the subjective feedback comparing with an HMD solution. The results of these two parts of the pilot study are presented as below.

4.2.1 Objective Evaluation

This experiment aimed to evaluate the starting and stopping latency of the WIPLS. According to the experiment results, the average starting latency is 287ms (standard deviation: 121ms); and the mean stopping latency was 781ms (standard deviation: 44ms). The longer stopping latency was due to the speed smoothing method; as mentioned above, to reduce the jerkiness caused by sudden changes in walking speed, the speed was smoothed by averaging the value in four frames. To overcome this issue, the speed smoothing method needed to be modified. The smoothed speed was set to zero if the speed before smoothing was zero. After this improvement, the stopping latency was reduced from 781ms to 474ms (standard deviation: 35ms). Thus, the mean starting latency (287ms) and mean stopping latency (474ms) in this research were under acceptable levels, as compared to the value reported in previous studies (500ms) (Usoh et al., 1999; Nilsson et al., 2013). Beside the latency, there was no apparent jerkiness reported by participants during the experiment.

Advantages of the ZCB algorithm included the lack of a calibration requirement and the ability to work with various body sizes. To evaluate these characteristics of the system, this research conducted a Mann-Whitney U test on the starting latency and the stopping latency for two subgroups' data (with significance level set at 0.05). Participants were divided into two groups by the median of the population height. One group included taller participants (8 participants, higher than 68 inches). The other group included shorter participants (9 participants, shorter than 68 inches). The null hypothesis was that there is no difference in the starting latency or stopping latency between two groups. Because the p-values of the test on starting latency and stopping latency were 0.1453 and 0.1181 respectively, it was unable to reject the null hypothesis. Therefore, it was safe to claim that the height of the participant does not affect the starting or stopping latency in the WIP system.

4.2.2 Subjective Survey Analysis

This experiment was conducted to collect subjective feedback from the participants on how they feel about the WIPLS comparing to the HMD solution. After collecting the response data from the survey questions, a Mann-Whitney U test was conducted on each question item. First, the basic statistics were compared. Table 4.1 shows a list of non-parametric statistics for two systems on each question item. From Table 4.1, it is able to find out that the HMD system was favored on most of the question items based on the participants' rating. These items included naturalness, responsiveness, immersion, and so on. This was as expected since the HMD can output stereoscopic image that provided more visual immersion to the users than other less immersive systems. Besides, the keyboard controlling interface had undoubtedly lower latency. From Question 3 and 4, it was able to find that the ZCB-WIP system resulted in less fatigue and motion sickness when compared to the HMD system. To determine whether those differences are significant, the Mann-Whitney U test was conducted on those two items. The hypotheses included:

- H_{a0} The WIPLS has equal or higher fatigue than the HMD system
- H_{a1} The WIPLS has lower fatigue than the HMD system
- H_{b0} The WIPLS has equal or higher motion sickness than the HMD system
- H_{b1} The WIPLS has lower motion sickness than the HMD system

Table 4.1: Basic quantile statistics of comparison between WIPLS and HMD VR systems

	ZCB-WIP	**HMD**
Q1: natural	[2.00, 2.00, 3.00]	[2.25, 4.00, 5.00]
Q2: responsive	[2.00, 2.00, 3.00]	[4.25, 5.00, 5.00]
Q3: fatigue	[1.00, 1.00, 2.00]	[1.00, 2.00, 3.00]
Q4: sickness	[1.00, 1.00, 1.00]	[1.00, 2.00, 4.00]
Q5: latency	[2.00, 3.00, 4.00]	[1.00, 1.00, 1.75]
Q6: immersion	[1.00, 2.00, 3.00]	[4.00, 5.00, 5.00]
Q7: easiness	[2.00, 3.00, 3.75]	[1.00, 2.00, 4.00]
Q8: comfortable	[2.00, 3.00, 3.00]	[3.00, 4.00, 5.00]

We performed a one-tailed Mann-Whitney U test on the two pairs of hypotheses and found that $p = 0.0432$ for H_a . Thus, it was able to reject the null hypothesis and conclude that the ZCB-WIP system results in lower fatigue than the HMD system. Similarly, since $p < 0.01$ for H_b, it was able to reject the null hypothesis and conclude that the WIPLS results in lower motion sickness than the HMD system. The conclusion that the VR with HMD causes more motion sickness and fatigue was consistent with prior studies (Kuze and Ukai, 2008; Martin et al., 2012; Moss et al., 2008).

4.3 Pilot Study 2: Survey Instrument Evaluation

According to Chapter 3, the 2nd pilot study was conducted to evaluate the survey instrument regarding reliability and validity.

The reliability of the survey items was the measurements on the overall consistency, and Cronbach's alpha is usually used to serve this purpose. According to previous literature, most of the research considered a coefficient higher than 0.70 to be acceptable. In this research, the survey instrument consisted of 12 Likert-scale questions measuring the perceived learning effectiveness and the satisfaction. The statistical analysis used the IBM SPSS Statistics Version 23 for the Cronbach's alpha calculation, and find out that the overall Cronbach's alpha for the data set of the pilot study was 0.799. This value was above the acceptable level, suggesting that the items had relatively high internal consistency. The detailed result of the reliability analysis is presented as below in Table 4.2. The last column of Table 4.2 is Cronbach's Alpha if an item is deleted, which shows how the overall Cronbach's alpha will change if the corresponding item is removed from the survey instruments. This coefficient is used to tell us which item is not highly correlated with other items. From the table, it is able to find out that only row SA1 and SA2 have a value greater than the original Cronbach's alpha, while the improved value after the deletion is very slight (from 0.799 to 0.801), thus these two items were kept in the survey instrument.

The reliability of a test alone is not sufficient; it also needs to be valid. Validity is an indicator of how well a test measures what it is purposed to measure. Since the items were selected out of the pool of published survey instruments (Chou and Liu, 2005; Lee, 2011), they had already been proven to have high construct validity. Further, feedback from the participants regarding their understanding of the questions was also collected in order to improve the clarity, appropriateness and readability. After several rounds of refining and rewording iteratively, the final version of the survey was confirmed, and no issues or misunderstandings were reported in this version.

Table 4.2: Results of reliability analysis on pilot study survey instruments

	Mean	Std. Deviation	N	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
LE1	4.54	1.374	28	0.376	0.791
LE2	4.14	1.604	28	0.485	0.78
LE3	5.25	1.295	28	0.502	0.78
LE4	4.32	1.492	28	0.499	0.779
LE5	5.46	1.427	28	0.609	0.768
LE6	4.96	1.453	28	0.363	0.792
LE7	5.07	1.215	28	0.436	0.786
LE8	5.18	1.701	28	0.517	0.777
SA1	5.71	1.384	28	0.262	0.801
SA2	5.04	1.261	28	0.24	0.801
SA3	5.5	1.503	28	0.499	0.779
SA4	5.21	1.228	28	0.535	0.777

4.4 Empirical Case Study

The empirical case study is aimed to solve the conceptual framework and answer the research questions. The statistical results and simple interpretations are conveyed in this section. More detailed discussions and implications based on these statistical results will be presented in Chapter 5.

4.4.1 Sample Descriptive Statistic

The sample population in this research included 240 college students from the LESSP in 2016. The demographic and background data were collected from all participants, and those questions were all mandatory so that there were no missing data. Demographic information includes the frequency data among the participants with regards to gender and age, as displayed in Table 4.3. The frequency table shows that 25.4% of the participants were female, and the rest, 74.6%, were male. Also, most of the participants (97.1%) were between 19 to 26 years old, among them, the major age group was between 19 and 22. Only a small portion of the participants were older than 26. None of the participants were below 19.

Table 4.3: Frequencies table of gender and age

		Frequency	Percent	Valid Percent	Cumulative Percent
Gender	Female	61	25.4	25.4	25.4
	Male	179	74.6	74.6	100
	Total	240	100	100	
Age	19-22	152	63.3	63.3	63.3
	23-26	81	33.8	33.8	97.1
	Over 26	7	2.9	2.9	100
	Total	240	100	100	

The background data included the prior experience of video game and prior knowledge on VR. This information was collected in the survey because it's commonly accepted that one's video gaming experience and VR knowledge may reflect his/her interest and acceptance towards VR as a learning tool, thus, affecting the final learning outcome in the experiment. From Table 4.4, it is able to find out that 25.8% of the participants do not play video games at all, 36.7% claim to play video games for 1 to 3 hours per week on average, and 20.4% play for 4 to 6 hours. So, the majority of the participants (82.9%) did not play a lot of video games (under 6 hours). The VR knowledge report showed that VR was rather popular and most of the participants had heard about it at least once. Only 2.5% of the participants claimed that they had never heard of it before. All other participants knew about VR or used VR before participating in the research. Note that there was a fourth option for this question item, which is "I am an expert on this topic", while no participant chose that option. This may be because all the participants were humble and did not want to recognize themselves as "expert", but it was also an indication that the whole sample population had limited prior VR experience.

4.4.2 Results of Exploratory Factor Analysis

With those item removal criteria mentioned in Chapter 3, it is now possible to run the factor analysis with all the survey items concerning the learning outcome using SPSS. The Pattern Matrix from the output is presented in Table 4.5.

Table 4.4: Frequencies table of background information

		Frequency	Percent	Valid Percent	Cumulative Percent
Gaming Experience	None	62	25.8	25.8	25.8
	1 to 3 hours	88	36.7	36.7	62.5
	4 to 6 hours	49	20.4	20.4	82.9
	7 to 9 hours	19	7.9	7.9	90.8
	10 hours or more	22	9.2	9.2	100
	Total	240	100	100	
VR Knowledge	Never heard of	6	2.5	2.5	2.5
	Know about it but never experienced one by myself	156	65	65	67.5
	Used VR a couple of times	78	32.5	32.5	100
	Total	240	100	100	

Table 4.5: Pattern Matrix of the factor analysis on the original survey items

	Factor	
	1	2
SA1		0.821
SA2		0.896
SA3		0.888
SA4		0.753
LE1	0.346	0.255
LE2	0.782	
LE3	0.853	
LE4	0.885	
LE5	0.472	0.277
LE6	0.774	
LE7	0.62	0.235
LE8	0.635	
Extraction Method: Principal Axis Factoring. Rotation Method: Promax with Kaiser Normalization. a. Rotation converged in 3 iterations.		

According to the item removal criteria, it is easy to find out that the factor structure illustrated from Table 4.5 is not very clean because there are cross-loading items LE1 and LE5 on both factors. Also, the differences between the primary loading and the secondary loading are both smaller than 0.2. The survey item LE7 also has cross-loading, while it is best to keep this item since the primary loading is more than 0.2 larger than the secondary loading.

To achieve better factor structure, these problematic survey items (LE1 and LE5) are removed. The updated Pattern Matrix can be found in Table 4.6.

Table 4.6: Pattern Matrix of the factor analysis after removing items

	Factor	
	1	2
SA1		0.820
SA2		0.883
SA3		0.895
SA4		0.755
LE2	0.725	
LE3	0.845	
LE4	0.892	
LE6	0.787	
LE7	0.620	0.245
LE8	0.648	
Extraction Method: Principal Axis Factoring. Rotation Method: Promax with Kaiser Normalization. a. Rotation converged in 3 iterations.		

Now it is able to find out that the factor structure is very clean, as the convergence and discriminant validity are evident with all survey items having high loadings on the factors. There is still one item with cross-loading on multiple factors, however. As the difference between the primary loading and the secondary loading are greater than 0.2, it can thus be considered as a valid item.

For determining the number of factors, there is a debate over multiple criteria on whether a factor is statistically important to be chosen. One of the most commonly used methods is to retain the factors with a large eigenvalue. (Kaiser, 1960) recommended that all factors with eigenvalues greater than 1 should be kept. In this study, there were two factors with eigenvalue greater than

1, as displayed in the Scree Plot in Figure 4.1. With the line $y = 1$ drawn on the graph, it is easy to identify that there are two points above this line, indicating that two factors have an eigenvalue greater than 1.

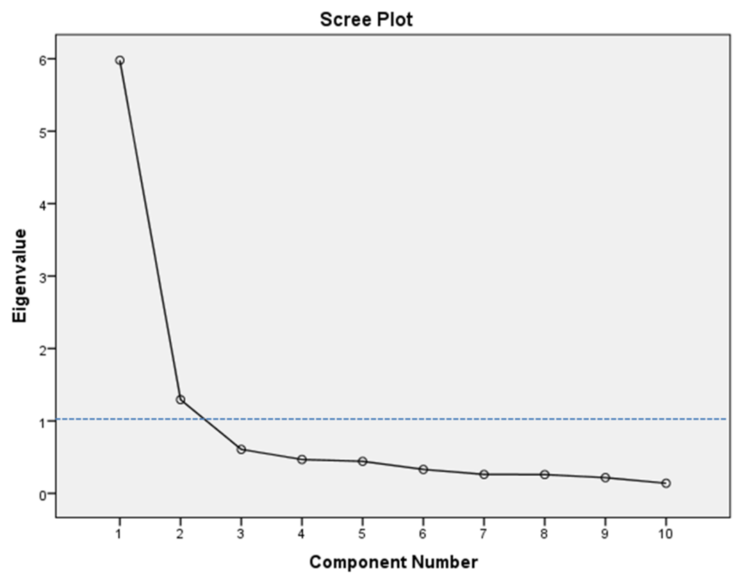

Figure 4.1: Scree Plot of the data with updated survey items

It is suggested that Scree Plot should not be the sole criterion for factor selection, and the total variances explained by the factors should also be examined. Table 4.7 from the output report of SPSS shows the amount of variance explained by the factors. As demonstrated in the table that the first two factors explained 72.733% of the total variance, indicating that these two factors are fairly adequate in representing the model.

Table 4.7: A Total variance explained by components

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared *Loadings*[a]
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total
1	5.979	59.792	59.792	5.979	59.792	59.792	5.304
2	1.294	12.94	72.733	1.294	12.94	72.733	4.849
3	0.607	6.074	78.806				
4	0.468	4.679	83.485				
5	0.442	4.423	87.908				
6	0.33	3.304	91.212				
7	0.262	2.62	93.832				
8	0.26	2.598	96.43				
9	0.218	2.175	98.605				
10	0.139	1.395	100				

Extraction Method: Principal Component Analysis.

a. When components are correlated, sums of squared loadings cannot be added to obtain a total variance.

The groups of items are also reasonable as all survey items associated with perceived learning effectiveness (LEs) were in one group, and all survey items associate with satisfaction (SAs) were loading on the other group.

Additionally, this research checked the KMO and Bartlett's test. From Table 4.8, it can be found that the value of KMO is 0.908 and the Bartlett's Test of Sphericity is significant, confirming that the data were appropriate for the factor analysis.

After the factor structure was explored in EFA, the CFA was conducted to impute the composite variables. Two factors, perceived learning effectiveness and satisfaction, were added as new columns in the data set, and these two variables instead of the raw survey items were used as the response variables in the remainder of this research.

Table 4.8: KMO and Bartlett's Test

KMO and Bartlett's Test		
Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		0.908
Bartlett's Test of Sphericity	Approx. Chi-Square	1718.477
	df	45
	Sig.	0

4.4.3 ANOVA after ART

With the ART procedure, it is possible to perform ANOVA on the nonparametric response variables obtained from the CFA. The response variables are perceived learning effectiveness (LE) and satisfaction (SA). The processing log from ARTool can be found in Figure 4.2.

Analysis of Perceived Learning Effectiveness

After processing the dataset VRData_LE.csv using ARTool, a new dataset with the file name of VRData_LE.art.csv was produced. The data structure of the original dataset, as well as the dataset after the ART processing, is displayed in Figure 4.3.

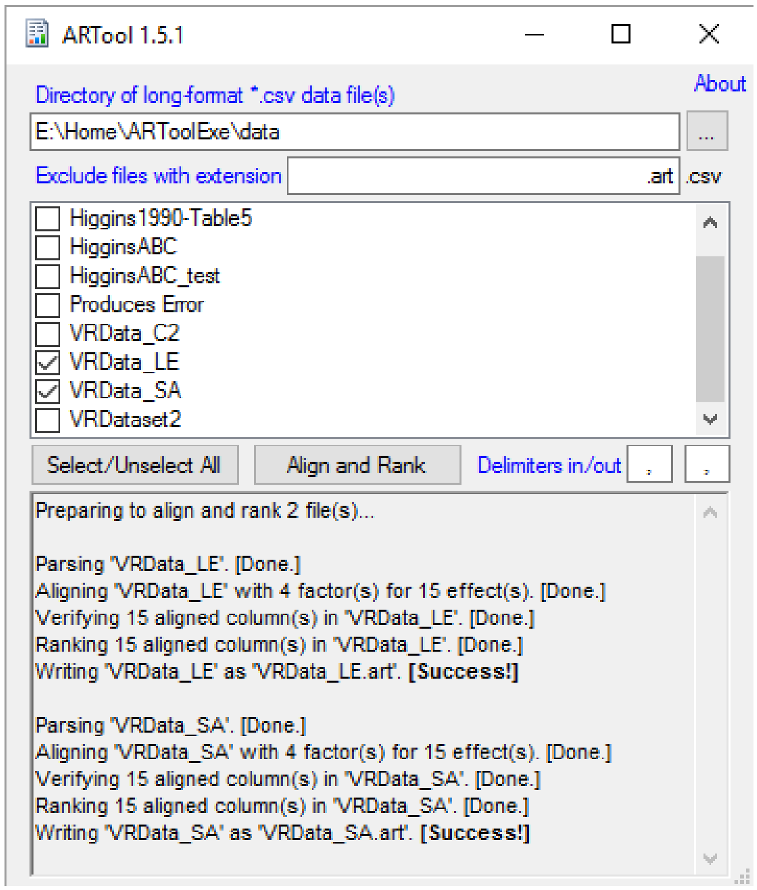

Figure 4.2: Using ARTool to process nonparametric data

As can be seen in Figure 4.3, only those highlighted columns are analyzed in this research, which is ranking data of the main effects and the two-degree interaction effects. The higher interaction effects are ignored in accordance with the design of experiment part of this research.

With those ten columns obtained from the ART procedure, this research now performs the ANOVA on each of these ranking response variables, respectively, and the corresponding ANOVA tables are listed in Appendix D.

From those candidate ANOVA tables, it is easy to identify that the significant term with a critical p-value of 0.05 is an interaction term X1Vi*X2NS (p-value = 0.039). Another term X2NS is also marginally significant (p-value = 0.069). All other terms are insignificant.

The next step was to interpret the results from the data to see if any insightful conclusions can be drawn. To better analyze the interaction, this research fitted a univariate model with only X1Vi, X2NS, and the interaction X1Vi*X2NS as the terms and the perceived learning effectiveness before the ART procedure as the dependent variable. Since the interaction term was significant, there was no point to interpret the main effects. The focus was put on interpreting the interaction term from this new model and the interaction plot. With X1Vi on horizontal axis and X2NS on separate lines, the interaction plot is now displayed in Figure 4.4. From the plot, it can found that when the visualization factor (X1Vi) is at low level, which is implemented in the WIPLS as black and white display, the higher level of natural semantics, i.e. using body language for system control, produces a higher perceived learning effectiveness than the less natural way of controlling, i.e. control using a traditional game controller. When the visualization factor is at the high level, which is implemented as a full-colored display, the difference between high level and low level of natural semantics is subtle, and in this experiment the less natural way of controlling slightly outperformed the more natural controlling mechanism. The detailed discussion and explanation of this result will be presented in Chapter 5.

The other factors in the model which were insignificant were Interaction (X3In) and Immersion (X4Im). This does not comply with intuition, since the factor Interaction (X3In) determined whether the VR system was a game-based environment or a video-based system, and the implementation details were fundamentally different between these two types of systems. Although it was expected that the game-based environment would outperform the video-based system, the

<table>
<tr><td colspan="3" rowspan="6">Columns in original data</td><td>Subject</td></tr>
<tr><td>X1Vi</td></tr>
<tr><td>X2NS</td></tr>
<tr><td>X3In</td></tr>
<tr><td>X4Im</td></tr>
<tr><td>LearnEff</td></tr>
<tr><td rowspan="30">Additional columns after processed by ARTool</td><td colspan="2" rowspan="15">Additional columns after alignment</td><td>aligned(LearnEff) for X1Vi</td></tr>
<tr><td>aligned(LearnEff) for X2NS</td></tr>
<tr><td>aligned(LearnEff) for X1Vi*X2NS</td></tr>
<tr><td>aligned(LearnEff) for X3In</td></tr>
<tr><td>aligned(LearnEff) for X1Vi*X3In</td></tr>
<tr><td>aligned(LearnEff) for X2NS*X3In</td></tr>
<tr><td>aligned(LearnEff) for X1Vi*X2NS*X3In</td></tr>
<tr><td>aligned(LearnEff) for X4Im</td></tr>
<tr><td>aligned(LearnEff) for X1Vi*X4Im</td></tr>
<tr><td>aligned(LearnEff) for X2NS*X4Im</td></tr>
<tr><td>aligned(LearnEff) for X1Vi*X2NS*X4Im</td></tr>
<tr><td>aligned(LearnEff) for X3In*X4Im</td></tr>
<tr><td>aligned(LearnEff) for X1Vi*X3In*X4Im</td></tr>
<tr><td>aligned(LearnEff) for X2NS*X3In*X4Im</td></tr>
<tr><td>aligned(LearnEff) for X1Vi*X2NS*X3In*X4Im</td></tr>
<tr><td rowspan="15">Additional columns after ranking</td><td rowspan="10">Columns of data used in this research</td><td>ART(LearnEff) for X1Vi</td></tr>
<tr><td>ART(LearnEff) for X2NS</td></tr>
<tr><td>ART(LearnEff) for X1Vi*X2NS</td></tr>
<tr><td>ART(LearnEff) for X3In</td></tr>
<tr><td>ART(LearnEff) for X1Vi*X3In</td></tr>
<tr><td>ART(LearnEff) for X2NS*X3In</td></tr>
<tr><td>ART(LearnEff) for X4Im</td></tr>
<tr><td>ART(LearnEff) for X3In*X4Im</td></tr>
<tr><td>ART(LearnEff) for X1Vi*X4Im</td></tr>
<tr><td>ART(LearnEff) for X2NS*X4Im</td></tr>
<tr><td rowspan="5">Columns of data ignored in this research</td><td>ART(LearnEff) for X1Vi*X2NS*X3In</td></tr>
<tr><td>ART(LearnEff) for X1Vi*X2NS*X4Im</td></tr>
<tr><td>ART(LearnEff) for X1Vi*X3In*X4Im</td></tr>
<tr><td>ART(LearnEff) for X2NS*X3In*X4Im</td></tr>
<tr><td>ART(LearnEff) for X1Vi*X2NS*X3In*X4Im</td></tr>
</table>

Figure 4.3: Data table structure of variable LearnEff after processing using ART

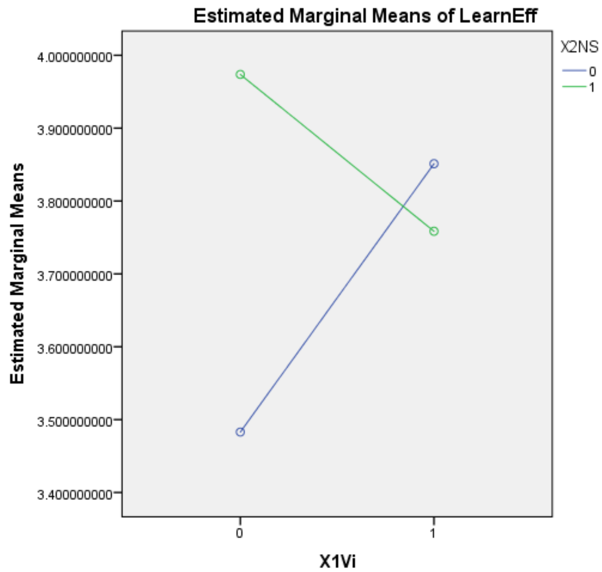

Figure 4.4: Interaction plot with X1Vi and X2NS

results of the analysis failed to show a significant difference. The factor Immersion (X4Im) also showed no significance, while this conclusion was based on the assumption that wider FOV (field of view) brings higher immersion (Duh et al., 2001; Prothero and Hoffman, 1995) . In this research, the VR system was implemented following this theory, with a wider FOV representing the high level of Immersion. So the insignificance of Immersion also means the insignificance of FOV in affecting the perceived learning effectiveness. These insignificant statistical results will also be discussed in detail in Chapter 5.

Analysis of Satisfaction

With the same ART procedure on the dataset VRData_SA.csv, the corresponding dataset with file name VRData_SA.art.csv was also obtained. The columns in this data table were similar with the VRData_LE.art.csv, with 30 additional columns, 10 of them being used as dependent variables in fitting the ANOVA model for each of the main effects and two-degree interaction terms. The Columns are displayed in Figure 4.5. The common ANOVA can now be performed on each of the response variables, respectively, and the corresponding ANOVA tables are listed in Appendix D.

Columns of data used in this research after ART procedure	ART(Satisfaction) for X1Vi
	ART(Satisfaction) for X2NS
	ART(Satisfaction) for X1Vi*X2NS
	ART(Satisfaction) for X3In
	ART(Satisfaction) for X1Vi*X3In
	ART(Satisfaction) for X2NS*X3In
	ART(Satisfaction) for X4Im
	ART(Satisfaction) for X3In*X4Im
	ART(Satisfaction) for X1Vi*X4Im
	ART(Satisfaction) for X2NS*X4Im

Figure 4.5: Interaction plot with X1Vi and X2NS

From those candidate ANOVA tables, the research can identify that the only significant term with a critical p-value of 0.05 was the main effect X2NS ($p - value = 0.032$). All other terms were insignificant.

To interpret the meaning of this significance, the marginal means of the response variable are plotted with all levels of the factor X2NS on the horizontal axis. From the plot in Figure 4.6, it is easy to find that the high level of Natural Semantics (X2NS), implemented as the game-based environment, will receive more satisfaction comparing to the low level, which is implemented as the video-based system.

4.4.4 Results of Group Analysis on Background Variables

The sample population is Levene's tests are also used to evaluate the homogeneity of the variance among those groups. This test should be insignificant to meet the assumption that the variances in each group were equal.

The results of the Leven's tests are presented in Table 4.9, Table 4.10 and Table 4.11.

Using 0.05 as the significance level, it was able to learn from the Levene's statistic that all variance homogeneity tests have p-value greater than 0.05. Thus, it was unable to reject the null

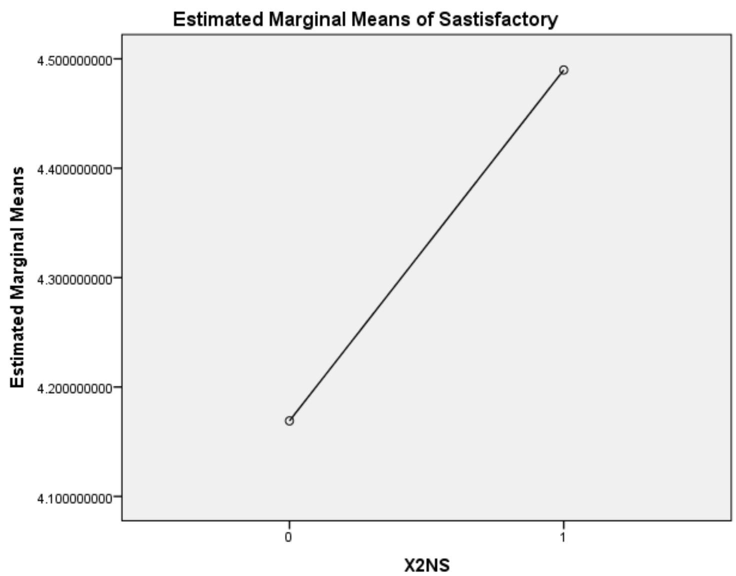

Figure 4.6: Main effects plot for satisfaction

Table 4.9: Test of homogeneity of variance between groups divided by gender

	Levene Statistic	df1	df2	Sig.
Learning Effectiveness	2.076	1	238	0.151
Satisfaction	3.84	1	238	0.051

Table 4.10: Test of homogeneity of variance between groups divided by video gaming experience

	Levene Statistic	df1	df2	Sig.
Learning Effectiveness	1.385	4	235	0.24
Satisfaction	1.189	4	235	0.316

Table 4.11: Test of homogeneity of variance between groups divided by prior VR experience

	Levene Statistic	df1	df2	Sig.
Learning Effectiveness	1.301	2	237	0.274
Satisfaction	0.011	2	237	0.989

hypothesis, and conclusion can be drawn that the assumption of equal variances among all the groups divided by the group variables was valid.

The results of the Kruskal-Wallis H test are displayed in Table 4.12, Table 4.13 and Table 4.14.

Table 4.12: Kruskal-Wallis Test on groups using gender as the grouping variables

	Learning Effectiveness	Satisfaction
Chi-Square	0.155	2.075
df	1	1
Asymp. Sig.	0.694	0.15

Table 4.13: AKruskal-Wallis Test on groups using video game experience as the grouping variables

	Learning Effectiveness	Satisfaction
Chi-Square	6.747	3.35
df	4	4
Asymp. Sig.	0.15	0.501

Table 4.14: Kruskal-Wallis Test on groups using prior VR experience as the grouping variables

	Learning Effectiveness	Satisfaction
Chi-Square	8.728	8.183
df	2	2
Asymp. Sig.	0.013	0.017

The result of the Kruskal-Wallis H tests showed that a significant difference between the groups of participants was found ($p < 0.05$). The test on prior VR experience revealed a significant difference in both percieved learning effectiveness and satisfaction among the groups with different prior VR experience, and the group with the higher level of prior VR experience showed a higher perceived learning effectiveness and satisfaction than the group with the lower level of prior VR experience. This may indicate that the participants with higher prior VR experience were more familiar with the VR technology. As a result, the benefit and attractive features of the VR applications in their previous experience made it easier for them to accept the VR technology in the educational field, and they were more willing to explore the learning content in any new virtual learning environment. Thus, their perception of percieved learning effectiveness as well as the satisfaction towards the VR learning environment would be higher. All other groups showed no significant differences.

4.4.5 Group Means Comparison for Hypotheses Tests

The statistics results of the Mann-Whitney U test of the hypotheses are listed in Appendix A. From the results table, it can be seen that the hypotheses that demonstrate significance are H_{1g}, H_{1h}, H_{1j} and H_{1l}. These hypotheses and the corresponding interpretations are as follows:

- H_{1g} Participants in VR subgroup with a high level of natural semantics will give a different rating on active learning than the VR subgroup with low level of natural semantics.
- H_{0g} Participants in VR subgroups with a high level and low level of natural semantics will give the same rating on active learning.

Interpretation: Since the $p - value = 0.016$, this research can reject the null hypothesis with the significance level of 0.05 and accept the alternative hypothesis that there is a significant difference between the groups with a high level and low level of natural semantics in the rating of active learning. This result of significance is within the expectation since the new way of controlling will produce novelty comparing to a traditional way of controlling, which made the participants eager to explore the virtual world and seek more learning content actively.

- H_{1h} Participants in VR subgroup with a high level of natural semantics will give a different rating on interactive learning than the VR subgroup with low level of natural semantics.
- H_{0h} Participants in VR subgroups with a high level and low level of natural semantics will give the same rating on interactive learning.

Interpretation: Since the $p - value = 0.018$, this research can reject the null hypothesis wit the significance level of 0.05 and accept the alternative hypothesis that there was a significant difference between the groups with a high level and low level of natural semantics in the rating of interactive learning. This result met the expectation, as the interactive learning was perceived through the interaction between the users and the system, and the interaction type that is more natural would inspire the users to explore the system further and seek feedback from the system to construct knowledge on their own.

- H_{1j} Participants in VR subgroup with a high level of natural semantics will give a different rating on control than the VR subgroup with low level of natural semantics.
- H_{0j} Participants in VR subgroups with a high level and low level of natural semantics will give the same rating on control.

Interpretation: The p-value in this hypotheses test is 0.021. Thus this research can reject the null hypotheses with the significance level of 0.05 and accept the alternative hypotheses that the rating of control in the group with a high level of natural semantics would be significantly different from the group with the low level of natural semantics. Because the high level of natural semantics was implemented with body language as the controlling mechanism comparing to the traditional game controller used by a low level of natural semantics, it was safe to conclude that the natural body language can provide more controllability over the traditional way. Moreover, this increased degree of controllability would result in a higher level of intrinsic motivation for learning.

- H_{1l} Participants in VR subgroup with a high level of natural semantics will give a different rating on experience than the VR subgroup with low level of natural semantics.
- H_{0l} Participants in VR subgroups with a high level and low level of natural semantics will give the same rating on experience.

Interpretation: Since the $p - value = 0.003$, this research can reject the null hypothesis with the significance level of 0.05 and accept the alternative hypothesis that there was a significant difference between the groups with a high level and low level of natural semantics in rating of experience. With a higher degree of natural semantics, the participants would have the chance to experience something different from conventional gaming experience. This novelty of experience would attract users to explore the virtual environment with more willingness. Thus, higher intrinsic motivation is also achieved.

Since four null hypotheses are rejected, the grand null hypotheses is also rejected:

- H_1 At least one hypothesis from H_{0a} to H_{0x} will be rejected.
- H_0 Hypotheses from H_{0a} to H_{0x} will all fail to be rejected.

Moreover, it can be inferred that there are significant differences in the ratings of the critical components between groups determined by the VR factors, and the VR factors are correlated with the theoretical learning frameworks.

4.4.6 Results of Model Selection Using Information Criteria

Instead of just using ANOVA to test the significance of our multivariate regression model that replies solely on p-value, this research also conducted the model selection approach for the research model using information criteria. An automated model selection using the *dredge* function in R was used, with AIC, AICc, BIC, Cp, and ICOMP as the information criteria. The results of the models with the various information criteria are displayed in Table 4.15 and 4.16. From Table 4.15, it can be seen that the model with lowest value of AIC, AICc and Cp was model 7. While the ICOMP was not the lowest, the difference was negligible (653.2 versus 653). So it is safe to conclude that the model 7 with X2NS and X3In was the optimal model according to the information criteria. This was consistent with the results of the ART approach, as they both considerred X2NS as the significant factor. Similarly, for the model using satisfaction as the response variable, the model 7 reported the lowest value of AIC, AICc, Cp and ICOMP, which indicating that this was the optimal model as well.

Table 4.15: Model selection using information criteria with perceived learning effectiveness as the response variable

	(Intrc)	X1Vi	X2NS	X3In	X4Im	AIC	BIC	Cp	ICOMP	AICc	df	logLik	delta	weight
7	3.558		0.199	0.2175		660	674	218	653.2	660.2	4	-326.016	0	0.185
5	3.658			0.2175		660.7	671.1	218.6	655.5	660.8	3	-327.35	0.6	0.137
3	3.667		0.199			661.2	671.7	219.1	656	661.3	3	-327.608	1.11	0.106
8	3.52	0.07647	0.199	0.2175		661.6	679	219.5	653	661.9	5	-325.818	1.69	0.079
1	3.767					661.8	668.8	219.7	657.8	661.9	2	-328.924	1.7	0.079
15	3.575		0.199	0.2175	-0.03303	662	679.4	219.8	653.3	662.2	5	-325.979	2.01	0.068
6	3.62	0.07647		0.2175		662.3	676.2	220.1	655.4	662.5	4	-327.154	2.27	0.059
13	3.674			0.2175	-0.03303	662.6	676.5	220.4	655.8	662.8	4	-327.313	2.59	0.051
4	3.629	0.07647	0.199			662.8	676.7	220.6	655.9	663	4	-327.412	2.79	0.046
11	3.684		0.199		-0.03303	663.1	677.1	220.9	656.3	663.3	4	-327.571	3.11	0.039
2	3.728	0.07647				663.5	673.9	221.2	658.3	663.6	3	-328.73	3.36	0.034
16	3.542	0.09841	0.199	0.2175	-0.06583	663.4	684.3	221.1	653	663.7	6	-325.687	3.53	0.032
9	3.783				-0.03303	663.8	674.2	221.4	658.6	663.9	3	-328.888	3.67	0.029
14	3.642	0.09841		0.2175	-0.06583	664	681.5	221.7	655.4	664.3	5	-327.024	4.1	0.024
12	3.651	0.09841	0.199		-0.06583	664.6	682	222.2	656	664.8	5	-327.283	4.62	0.018
10	3.75	0.09841			-0.06583	665.2	679.1	222.8	658.3	665.4	4	-328.603	5.17	0.014

Table 4.16: Model selection using information criteria with satisfaction as the response variable

	(Intrc)	X1Vi	X2NS	X3In	X4Im	AIC	BIC	Cp	ICOMP	AICc	df	logLik	delta	weight
7	4.063		0.3207	0.2129		721.4	735.3	281.5	714.5	721.6	4	-356.694	0	0.234
3	4.169		0.3207			721.8	732.2	282	716.6	721.9	3	-357.877	0.3	0.202
8	4.029	0.06773	0.3207	0.2129		723.1	740.6	283.6	714.5	723.4	5	-356.574	1.85	0.093
15	4.078		0.3207	0.2129	-0.03098	723.3	740.7	283.8	714.7	723.6	5	-356.669	2.04	0.085
4	4.135	0.06773	0.3207			723.5	737.4	284	716.6	723.7	4	-357.758	2.13	0.081
11	4.185		0.3207		-0.03098	723.7	737.6	284.3	716.8	723.9	4	-357.852	2.32	0.074
5	4.223			0.2129		724.7	735.2	285.5	719.5	724.8	3	-359.362	3.27	0.046
1	4.329					725	732	285.8	721	725.1	2	-360.519	3.53	0.04
16	4.049	0.08781	0.3207	0.2129	-0.06025	725	745.9	285.8	714.6	725.3	6	-356.489	3.78	0.035
12	4.155	0.08781	0.3207		-0.06025	725.3	742.8	286.2	716.7	725.6	5	-357.674	4.05	0.031
6	4.189	0.06773		0.2129		726.5	740.4	287.6	719.6	726.7	4	-359.244	5.1	0.018
13	4.238			0.2129	-0.03098	726.7	740.6	287.8	719.8	726.8	4	-359.337	5.29	0.017
2	4.296	0.06773				726.8	737.2	287.9	721.6	726.9	3	-360.402	5.35	0.016
9	4.345				-0.03098	727	737.4	288.2	721.8	727.1	3	-360.495	5.53	0.015
14	4.209	0.08781		0.2129	-0.06025	728.3	745.7	289.8	719.7	728.6	5	-359.161	7.02	0.007
10	4.316	0.08781			-0.06025	728.6	742.6	290.2	721.8	728.8	4	-360.32	7.25	0.006

4.5 Summary

In this chapter, research findings based on the survey data were explored and presented. The survey data were examined and analyzed to identify which VR factors were impacting the learning outcome of WIPLS. The exploratory factor analysis and confirmative factor analysis were used to remove some survey items that were not internally consistent with the overall measured variables and collapse the groups of survey items into composite variables: perceived learning effectiveness and satisfaction. While these composite variables were still nonparametric, to perform the traditional ANOVA procedure and explore the significant main effects and interaction effects, this research used the ART procedure to transform and rank the response variables. ART procedure generated additional data columns as response variables and traditional ANOVA could be fit to identify the significant terms. The interpretations of the findings of main effects as well as interactions were also performed. Group mean analyses were conducted on the ratings of critical components of theoretical learning frameworks, using the VR factors as grouping variables in dividing the sample population into pairs of subgroups. This research used the Mann-Whitney U test for these nonparametric unpaired group mean tests. The results from the group mean analysis rejected some of the hypotheses, thus, proving the correlations between the VR factors and the theoretical learning frameworks such as constructivist-based learning approach and the intrinsic motivation to learning.

The purpose of this chapter was to analyze the data from the sample population using appropriate statistic procedures and report the results with interpretation. Those results will be discussed in detail in the next chapter.

Chapter 5

Discussions and Implications

5.1 Overview

The aim of this study was to find out "how VR impact the learning outcome of the learners". It presented a WIPLS system with customizable VR factors and used a DOE approach that manipulated the levels of those VR factors to analyze how the learning outcome was determined. A research model exploring the relationship between the VR factors and the critical components of theoretical learning components was also proposed.

This research conducted experiments on a sample population of 240 participants and collected their response data using a survey. The survey items included the perceived learning effectiveness and the satisfaction as well as the ratings of the critical components of the constructivist-based learning approach and the intrinsic motivation theory. Appropriate statistical methods were applied to explore any statistical results behind those experiment data.

This chapter presents the discussions and implications obtained from the empirical results. In the previous chapter, some simple interpretations were provided following the statistical results. This chapter explains those statistical results in more depth. Instead of making the statements on the surface, it takes one step ahead and explores the theory behind the statistical output. Also, this research compares the conclusions and implications of the research with the findings from the existing literature and explains any compliance as well as contradictions from those comparisons. Additionally, this research tries to answer the “so what” question, and provides insightful suggestions and recommendations as well as caveats for future VR practitioners in their development of educational VR systems.

5.2 Significant VR factor: Natural Semantics

The results of ANOVA analysis on the perceived learning effectiveness and satisfaction as well as the group mean comparison on the critical components showed something in common: that the natural interaction was the most significant factor making the differences. Natural semantics is defined as the manner of behavior that is intuitive and natural, with the objective to minimize the burden of learning new knowledge and make use of what the users already know (Winn et al., 1993). In the WIPLS, natural semantics was implemented as using the natural walking and turning behaviors as the control mechanism to navigate the movement in the virtual world. Since walking and turning were in most people’s basic skill sets, there was no need for them to spare any extra effort learning them. On the contrary, in the VR treatments with low level of natural semantics, the participants had to use a game controller for similar navigational movements, and a mapping from “press a button” to “walk ahead in the virtual world” needed to be established, yielding a less natural way of interacting with the system.

The natural semantics is a concept that has been widely adopted in the VR field as well as traditional video game industry. For example, the Nintendo Wii is a game console that allows the users to play various video games, especially sports related games, using the handheld controller like swinging a racket and strike the virtual ball in the gaming environment. There is minimal instruction or practice required for the players if they have similar sports experience in the real world.

Insights can be provided to the VR practitioner on how to make use of natural semantics. For educational VR systems with the purpose of gaining hands-on experiences and skills that require practice, the natural semantics may play a crucial role. For the traditional teaching paradigm, knowledge and expertise are transferred to the learners in highly abstract forms through static media, like textbooks, lectures, and quizzes. In some literature, these abstracted forms are called symbol systems (Winn et al., 1993). The learners will need to perform a translation process to convert these abstract forms into concrete forms that are comprehensible and easy to follow so that they can practice and finally acquire it. This translation from abstract form to concrete form is indirect and adverse to the learning outcome because it is artificial and requires extra cognition effort to build a mapping relationship between the abstract symbols and the concrete substance. Unfortunately, this abstract form of knowledge is indispensable in traditional teaching paradigms because only the abstract form can be stored in traditional media. This is not the case if using VR as the new media. As displayed in Figure 5.1, in a VR learning system with a high level of natural semantics, the mapping from an abstract form of knowledge into a concrete form is no longer necessary since the users can access the concrete form of knowledge directly through a first-person experience. The learners can see and hear as well as feel in the same way as if they are in an authentic scenario and acquire the knowledge and skills directly. When they respond to the system accordingly in a natural way, instant feedback will also be provided to them. No more translation process is required before accessing the concrete form of knowledge and skills, and the information lost during these translation processes can be significantly reduced.

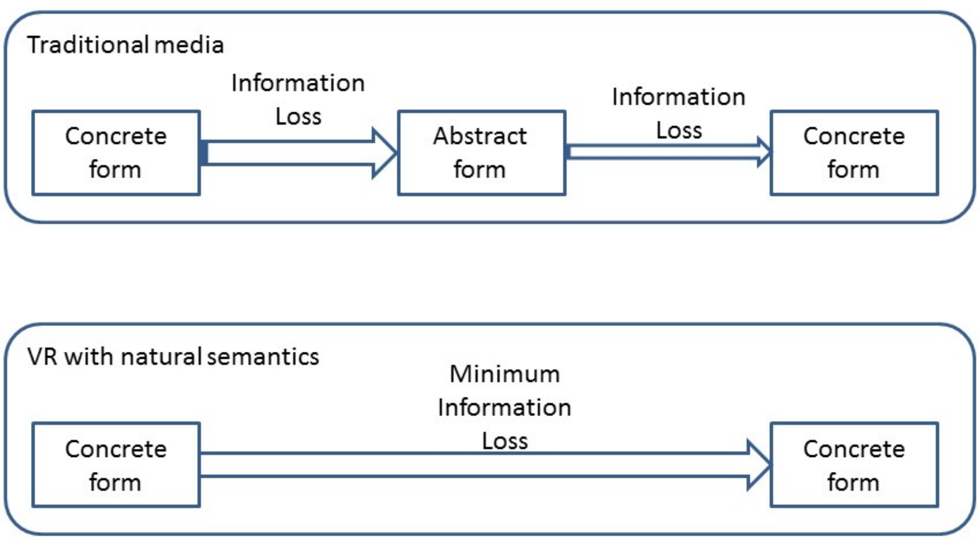

Figure 5.1: Knowledge transfer process with and without abstract form

Another great benefit of natural semantics is the support for constructivist-based learning approach, which has already been proven in the hypotheses tests on the relationship between the

VR factors and the critical components (H_{1g}, H_{1h}). The constructivists believe that learning is most effective and efficient if knowledge can be constructed by the learners themselves based on what they already knew in their prior experiences. Natural semantics provides a way of learning in which the learners can start the learning tasks quickly with short or no training time, since all prerequisites, if any, should be within the learners' existing skill sets. Learners can focus on the actual learning subjects in the learning activities and carry no burden on anything that is non-value-added to the learning objective.

5.3 Significant Interaction: Natural Semantics and Visualization

As can be seen from the results of the ANOVA on perceived learning effectiveness, there is a significant interaction between the Natural Semantics and Visualization, which is demonstrated in Figure 4.4. When the visualization is at a low-level, which is implemented as black and white in the WIPLS, the high level of natural semantics, implemented with body language as controlling mechanism, will produce a higher perceived learning effectiveness than the low level of natural semantics, which is implemented with a traditional game controller as the controlling method. However, when the visualization is at a high level, which is implemented as full color mode, there is only a marginal difference between the VR systems with a low level and high level of natural semantics. In other words, the advantage of a more natural VR failed to be observed under the full color condition.

This result seems to contradict with the intuition that the higher level of visualization will perform at least the same, if not better than the low level of visualization in terms of perceived learning effectiveness because people will always prefer a display with full color to the black and white display. Why, in the results, are the participants showing significant difference between the low level and high level of natural semantics only under the condition of low level of visualization instead of the high level?

The result can be explained by the malleable attentional resource theory. The malleable attentional resource theory presumes that there is a single pool of attentional resources that is shared among multiple tasks, and those tasks will compete for this pool of attentional resources, which is called competitive selection process.

In this research, the attention of the VR participants was distributed uniformly into multiple aspects, like the visual display, the controlling mechanism, and the feedback of the VR system. Since the total amount of attention was fixed for every individual, if one aspect of the VR took too much attention, the remaining attention available to other aspects were limited. For those participants experiencing the VR systems with a low level of visualization, since the display was presented in black and white, the participants only needed to spend a small portion of attention to process the colorless display and spent the majority of the attention on other aspects of the VR system. Thus, they would also have more attention on how the system was controlled, and it was for them to distinguish the differences between the high level and low level of natural semantics. On the contrary, for those participants experiencing the VR systems with a high level of visualization, a full-colored display was presented, providing a virtual world that was more vivid and closer to real life. The enriched virtual environment with high level of visualization was full of virtual objects, which kept the participants more engaged, while at the same time occupying more attention. Under these circumstances, the attentional resources remained for experiencing the natural semantics was proportionally reduced, and as a result, the participants were distracted and failed to report the difference in perceived learning effectiveness produced by the natural semantics.

Note that in this research, two response variables were measured: perceived learning effectiveness and satisfaction. While only the first response variable showed a significant difference in the interaction, there was no significant interaction between visualization and natural semantics in terms of satisfaction. The satisfaction was to measure the subjective feelings towards the VR system, with no emphasis on the learning subject, thus, it was less sensitive to the changes of attentional resources. Looking at the satisfaction results in the experiment, regardless of the colorless or full-colored display, the VR systems with body language as the controlling mechanism were preferred by the participants over the traditional controlling method. This supported the malleable attentional resources theory from another direction that if the attentional resources were less desired, the competition among multiple tasks would also decrease. The malleable attentional resources theory in this research is demonstrated in Figure 5.2.

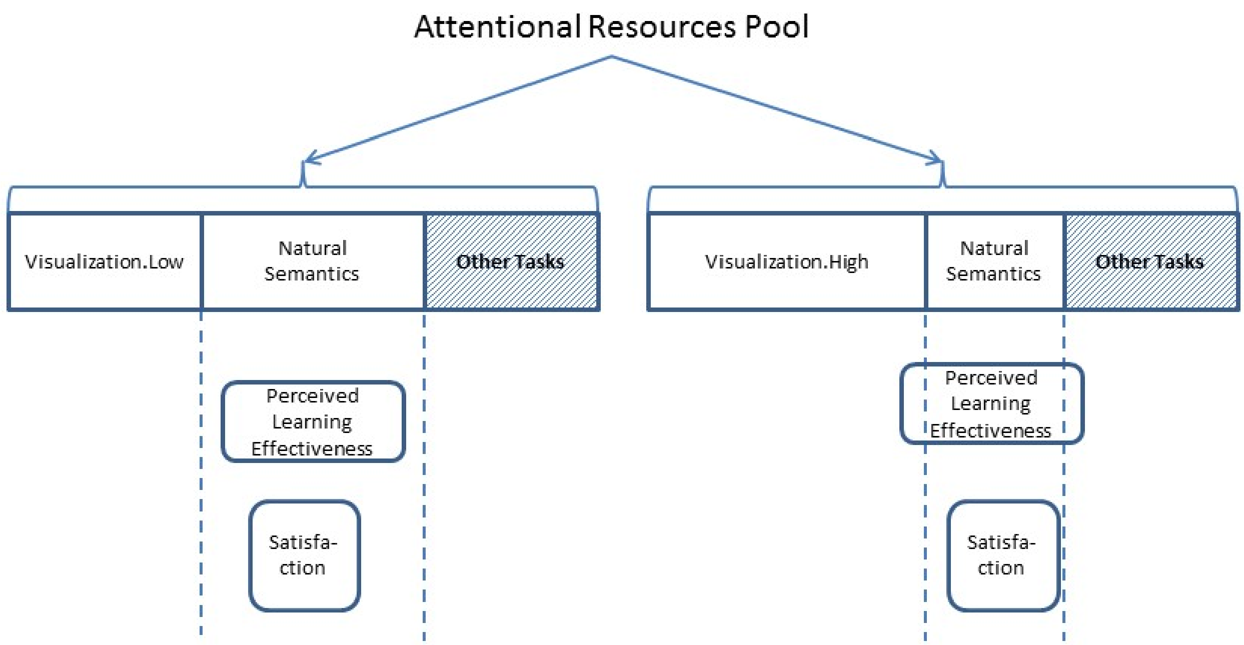

Figure 5.2: Interaction between visualization and natural semantics explained in malleable attentional resources theory

This malleable attentional resource theory can provide practical insights to VR practitioners for their future VR design and development. It is widely believed that in the field of VR, the improvement on visual experience will always result in "better" performance. Here the visual experience refers to all kinds of visual stimuli that can provide a better personal experience to the users, like higher resolution, higher color depth, and more polygons in the 3-D model. This is probably true for most of the cases, while exceptions may be raised if dealing with an educational VR system. According to the malleable attentional resource theory, the fixed amount of attentional resource needs to be divided among multiple tasks, including the learning content inside the VR system and the VR system itself. For such a VR system, an excessive amount of visual experience placed upon the learners may distract them from the learning subject and result in an adverse impact on their learning activities. One possible outcome of an educational VR system with enriched visual experience and gaming characteristics may be that the learners think the VR system is fun and appealing, while they gain little knowledge and skills out of the educational aspect of the VR system as they put too little attentional resources on it. Of course, this is not to say visual experience is useless, since an educational VR system that provides a low-level visual experience may make the learners lose interest in the whole VR system, and apparently no learning outcome will be achieved either. The VR practitioners should plan ahead and find the appropriate amount of visual experience that can attract the learners while not occupying too many attentional resources.

5.4 Insignificant VR Factor: Interaction

Interaction in this research is a factor that determines the amount of controllability the users can input into the system as well as the amount of feedback they can obtain from the system. In the WIPLS, the low level of interaction was implemented as a pre-recorded video that can only respond to simple user commands like "pause" or "play". The high level of interaction was implemented with more gaming characteristic, which took more complex commands and provided users with more responsive feedback.

The results of the ANOVA test and the hypotheses tests suggested that the interaction factor had no significant effect on the learning outcome. This indicated that the gaming characteristics may not necessarily produce a higher level of learning outcome. One possible reason is that not everyone prefers a gaming environment to a video environment, especially for those who do not have much interest or experience in a video game before. The survey data in Table 4.4 shows that most of the participants (62.5%) played video games less than 3 hours per week, which supports this theory. Also, the gaming characteristics need to be implemented in a way that not only attracts users but also matches the learning subject; otherwise the learning outcome achieved out of the virtual learning system might be compromised.

Another explanation of the insignificance might be that the low level of the interaction implemented using the video-based learning is not as ineffective as some VR practitioners thought. Although the video-based learning system lacks many advanced features comparing to the VR-based learning system, this method of learning has existed for several decades and is well accepted by the majority of the population nowadays.

5.5 Insignificant VR Factor: Immersion

Immersion is one of the most distinguishable factors that differentiate the VR from the traditional video games; it is defined as a subjective perception of being physically present in the virtual world. A positive correlation between the level of immersion and the learning effectiveness has been reported in several literary texts (Bangay and Preston, 1998; Pausch et al., 1997; Psotka, 1995; Vora et al., 2002), while this research did not found the same results. This discrepancy can be explained by examining the differences among research that reports a significant effect on immersion and that which does not.

Most of the literature reporting a significant effect on immersion was using head-mounted display (HMD) as the display device (Pausch et al., 1997; Psotka, 1995; Vora et al., 2002), while for non-immersive VR that used traditional screen as the displaying device, the effect of immersion was rarely reported (Burigat and Chittaro, 2007; Rahim and Eliana, 2013; Ryan et al., 2006). In this research, the participants were experiencing the virtual world through a flat screen. Although the screen size is much larger than the traditional computer monitors, there is always a noticeable boundary that distinguishes the real world from the virtual world. This sets an upper bound on the amount of immersion one can experience from the VR system. No matter how much the FOV was changed, as long as it used the non-immersive screen as the display output, it made no big difference on the amount of immersion the VR can provide. As shown in Figure 5.4, with the fixed-size screen, a wider angle of FOV displays more content on the screen, while at the same time, the entirety of the content are farther from the camera and look smaller when projected on the screen than in the narrower angle of FOV. A negative experience would be brought to the users when everything looks zoomed out, and the benefits brought by the wider angle of FOV would also be counterbalanced. This explained why a higher level of FOV did not bring a higher level of immersion as well as learning outcome ratings in the WIPLS.

This is not the case when using an HMD as the display device. Figure 5.4 shows the demonstration of what the users can see when wearing a cardboard style VR HMD. It can be found out easily how the FOV can affect the view of the users. Similar with displaying on a screen, more content is included in the view when a wider angle of FOV is used, while differences are that

Figure 5.3: Display of virtual world with narrower FOV (top) and wider FOV (bottom) in screen

the virtual world is not zoomed out, and the virtual objects look the same size as in the narrower angle of FOV. Obviously, when using HMD for displaying, the wider angle of FOV will always bring a higher level of immersion as well as better experiences.

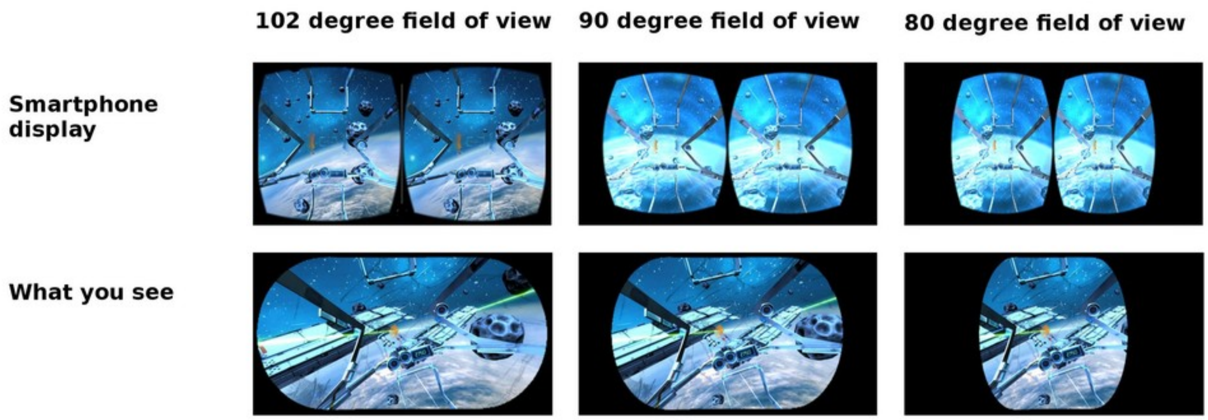

Figure 5.4: Display of virtual world with different FOV in HMD (Korolov, 2016)

Suggestions can be provided to the VR practitioners from the insignificant results in this research. First, the VR practitioners should consider whether or not as well as how much amount of FOV they should put into the VR application, depending on what type of VR is to be developed. If the HMD is chosen as the display device, they may want to implement it with a wide angle of FOV, so as to achieve the maximum amount of immersion. Instead, if the screen display is chosen as the display device, the FOV does not matter as much anymore, and an appropriate angle of FOV that can achieve a balance between the zoom levels and the amount of virtual objects in sight should be pursued. The correct amount of FOV should also be affected by the type of the VR application. If the VR is emphasized by integrating more information so as to make comprehensive decisions and reactions, e.g. a first-person shooting game, a wider angle of FOV may be helpful. However, if the VR focuses on the details of the virtual objects or the screen size is limited, e.g. an escape room game, a narrower angle of FOV may be useful. What's more, increasing FOV will always result in higher cost, which may cause a drop of frame per second (FPS) and bring a jerky experience if the hardware resource is limited. In such cases, properly lowering the FOV may be a good option so as to run the VR application smoothly while not overly sacrificing the immersion.

Another suggestion to the VR practitioner may be that immersion may be taken into consideration, depending on the purpose of the VR application. For a VR with learning outcome as its main objective, the level of immersion may not be a very important factor, since the tasks of acquiring new knowledge and skills are treated with higher priorities, and the subjective feelings to the VR system itself may be relatively diminished. This is similar to the serious games (Zyda, 2005) which emphasizes the pedagogical value and less care in if the game is fun and interesting to the players. However, if the VR application is more entertainment-oriented, the weight of immersion may be much higher, and the VR practitioners should consider taking the immersion as one of their main objectives during the designing stage of the development cycle.

5.6 Other Insights and Discussions

Besides using VR technology as the learning tools, other related applications can also take advantage of the conclusions from the results of this research. For example, instead of academic learning tasks among college students, some manufacturing tasks that requires extensive practice can be provided in the virtual environment. When practicing the tasks using the VR technology, the employees will have the chance to explore the outcomes of their decisions repeatedly without undertaking any risks on themselves or the equipments (Mujber et al., 2004). The interaction feature of the VR technology can also be beneficial in training the operators on complex and expensive machine, which will reduce the production cost and the training duration. This wil eventually the competency level of the employees (Olive et al., 2006).

Other fields that expect to gain dramatically from the VR technology are healthcare and rehabilitation. Comparing to traditional motion analysis from a video form data in rehabilitation applications, VR technology can provide a real-time and intelligent data analysis approach by making use of the motion sensors (Rizzo and Kim, 2005). This would greatly increase the rehabilitation efficiency. The inadequancy of training programmes for healthcare workers is also a severe issue. While it is widely believed that by offering VR based learning programmes for pre- and post-registration health professional education, invaluable educational experience can be achieved by the healthcare professionals in a cost effective and time effective manner (Saxena et al., 2016).

5.7 Summary

This chapter presented the discussion and implications of the research findings discovered in the previous chapter, including the interpretation, explanations, and theories on the significant VR factors, the significant interactions, and the insignificant VR factors. It explained the results by looking at the definition and making comparisons between the implementations in this study and the previous literature reporting similar or contradictory results. Suggestions and insights to the VR practitioners for their VR application development were also presented.

From the statistics results revealed in Chapter 4, this research found that some VR factors were more significant than the others in impacting the perceived learning effectiveness and the satisfaction. The natural semantics was the most significant VR factor in all the statistic tests performed in this research. This research concluded that the natural way of interacting with the system was more efficient and effective than the artificial approach because there was less information lost during the transferring of knowledge. The interaction between the natural interaction and the visualization was explained using the malleable attentional resource theory. Corresponding suggestions were provided to the VR practitioners on how to make use of these theories when designing and developing VR applications in the future.

Besides the significant effects, this research also explored the insignificant effects as well as the possible reasons behind those insignificant results. It then compared the results of this research with the literature reporting discrepant results and investigated the differences between them. This led to some insightful conclusions, which might be helpful in guiding the VR practitioners on how to invest the resources on some VR factors, depending on the type of VR applications they are planning to develop. One useful piece of information is that if the VR application is using HMD as the display device, immersion may play a major role; otherwise, the impact of immersion may be limited. Moreover, the purpose of the VR application is also a decisive factor on whether to emphasize on the immersion or not.

Chapter 6

Conclusions

6.1 Research Overview

This research investigated how the semi-immersive VR can be used as a learning tool by exploring the impact of the VR factors and the interactions between those factors. Theoretical learning frameworks including the constructivist-based learning approach and intrinsic motivation were also discussed, and hypotheses were proposed and tested to reveal the correlations between the VR factors and the critical components of those learning frameworks. A theoretical research model was developed to show how these variables are affecting each other.

The research was carried out in the following sequence:

First, a Walk-in-Place Learning System (WIPLS) was developed to provide a VR system with low latency, low jerkiness, and free of burden (Hongbiao Yang, 2015). What's more, the WIPLS is highly customizable and can generate a list of sub VR systems that are similar to each other while differing only in one or more VR factors. Experiments were also conducted to validate the WIPLS with the participation of the graduate students from the University of Tennessee. The objective performance and the subjective feedback were both evaluated. From the experiment results, conclusions can be drawn that the WIPLS is a well-developed VR system with great user experience.

Next, this research designed a survey instrument by referring the well-tested VR related questionnaires from previous literature and made necessary changes so that they can better fit the research purpose. The survey instrument included three groups of question items. The first two groups of question items measured the perceived learning effectiveness and the satisfaction of the WIPLS, while the last group measured the correlation between the VR factors and the critical components of the theoretical learning frameworks. A pilot study was conducted on participants recruited from Kids U summer camp at the University of Tennessee to test the internal consistency and the construct validity of the questionnaire. The results of the pilot study showed that the survey instrument was reliable and valid.

After evaluating the WIPLS system and the survey items, this research conducted the final experiment on the sample population of 240 participants from LESSP program at the University of Tennessee. Design of Experiment was used to generate the fractional factorial design with the treatment combinations of four VR factors, and those designed treatments were then implemented into the specific treatment sub-VR systems. There were in total 12 treatments in the fractional factorial design. Thus, it was able to test the main effects as well as the two-degree interactions between the VR factors. The last group of the question items was served to test the hypotheses and evaluate the correlations between the VR factors and the critical components of the theoretical learning frameworks. Based on the results obtained from the experiments, it was able to gain meaningful conclusions, as well as providing suggestions and insights to the VR practitioners.

The first two parts of this study were used to validate the research tools used in this study, and the last part was to answer the research question: How can VR be used as a learning tool, and which VR factor(s) is impacting the learning outcome?

From the experiment results, it is easy to find that natural semantics is the most significant factor that impacts satisfaction, and also supports the constructivist-based learning approach and the intrinsic motivation. This implies that increasing the naturalness of the VR system will be beneficial to the learning outcome because comparing to the traditional way of learning that uses traditional media to store knowledge, the VR-based learning system utilizes a more natural media that can reduce the information lost during the knowledge transfer from abstract form into concrete form.

There was a significant impact on the perceived learning effectiveness of the interaction between natural semantics and visualization. The results showed that when visualization was at the high level, the higher natural semantics led to a higher level of perceived learning effectiveness than the low level of natural semantics, while when the visualization was at the low level, the difference of perceived learning effectiveness between the low level and high level of natural semantics was insignificant. This result was interpreted with the malleable attentional resource theory. This theory assumes that the simultaneous tasks performed by an individual at the same time would compete for the same pool of attentional resources, and that one task taking too many attentional resources would result in insufficient attentional resources for the rest of the tasks. Since the participants were spending too much attentional resources on the VR with higher visualization, the attentional resources left for the natural semantics were limited, thus they were unable to focus on this feature and also unable to differentiate between the levels.

The insignificant VR factors, interaction and immersion, can be interpreted as follows. The interaction represents the gaming feature of the VR system. While the insignificance from this factor implies that the traditional video-based learning approach may still be a very effective learning approach, it will not be replaced by the VR-based learning approach in the near future. VR-based education should be served as a supplement to the traditional learning methods instead of a replacement. Also, the previous gaming experience of the sample population may also affect the results. Since in the sample population, most of the participants reported a relatively insufficient experience in video game playing, this might be one of the reasons that gaming characteristics in this research did not attract too much attention from the participants. Immersion was another VR factor that did not show any significance in impacting the learning outcome. This result was in contradiction with previous literature. After looking into the previous literature and the experiment conditions, it has been found that most of the literature reporting a significant impact in immersion was using HMD as the display device. However, in this research, a flat screen was used as the display apparatus. This may be the reason for this discrepancy since higher FOV in an HMD would make a tremendous impact on the users, while on a flat screen with fixed size, the FOV does not matter that much.

6.2 Contributions

This research 1) revealed how the factors of a VR system could impact the learning outcome by designing a list of comparable sub-VR systems, and 2) conducted an empirical study to validate the model so as to extract meaningful conclusions. There were many researchers that used VR for educational purposes and drew conclusions that VR is a beneficial tool in promoting learning outcome, although few researchers had investigated the VR system and explored how this objective is achieved. Most of the time, VR was just treated as a black box, without any knowledge of the internal mechanism. The researchers only created individual VR applications with their own design and conducted empirical studies by comparing with traditional educational approaches, while the exploration into the VR was impossible with just one individual VR application. Also, it was meaningless to make horizontal comparisons among different VR applications from previous literature to uncover their differences, since most of the VR applications were implemented heterogeneously when no universal metric can be used to evaluate these VR systems and make a fair comparison. Too much variation exists from one VR implementation to another to interpret the differences found among those VR systems. This issue can be handled by controlling all the uncounted variances and manipulating the variables of interest. This research developed a VR system named WIPLS to achieve this purpose. The WIPLS is a highly customizable VR application that can generate a pool of sub-VR systems that share most of the characteristics, while varying only in one or two variables at a time. In this way, everything is unchanged except the VR factors of interest, thus enabling the fair comparisons from one sub-VR application to another. As a result, any significant differences between the response variables can be attributed to the manipulated factors.

The WIPLS, as a highly customizable VR system is a contribution to the VR community. In this research, four commonly used VR factors were chosen to produce a pool of comparable sub-VR systems for analysis, while VR factors other than those four can also be incorporated into the WIPLS for extended research, as long as that VR factor can be implemented in the WIPLS. This enables us to test a broad range of VR factors within the limit of the WIPLS's capability.

The statistics results indicating the significant and insignificant VR factors and the interactions can provide meaningful insights and conclusions to the VR practitioners for their design and development of VR applications, which is another contribution of this research. When a VR practitioner is planning to develop a new VR application, there are ample design choices for him to make, like whether to go HMD or use the flat screen as the display device, how much resolution would the display need, or what control mechanism should the VR application use, etc. The VR practitioners can refer to this research when they face such issues and look for insights on how to implement each factor in their development of VR application and reach an appropriate solution that is within their budget limit, while at the same time maximizing their objective of the VR application.

Many empirical studies reported VR to be beneficial to the learning outcome and tried to explain their findings by looking at specific facts from their case study, without providing the explanation from the higher level (Coles et al., 2007; Mann et al., 2002; Vera et al., 2005; Vogel et al., 2006; Vora et al., 2002). This research proposed a theory that VR can promote the learning outcome through the theoretical learning frameworks from a higher level of perspective. According to this theory, VR can support the constructivist-based learning and increase the intrinsic motivation of the learners, thus, learning outcome can also be achieved. This research used data from the empirical study and looked for correlations between the VR factors and the critical components of these two theoretical learning frameworks. The results from the experiment supported this theory. This contribution can be used to provide theory support on using VR application as the learning tools.

6.3 Limitations

This research has several limitations that might affect the generalization of the conclusions:

First, the participants in this research are college students from the LESSP group, which might bring some bias to this study. The conclusions may not be able to be generalized into a wider range of the population. The problem of convenience sampling may also exist in this research.

Second, due to time limitation, the survey instrument was designed to include only a limited amount of question items to increase response rate. More tests can be done, and the variation can be further controlled if more time is allowed and more question items can be added to the survey instrument.

Third, the participants were attending the experiments in batch after their mid-term presentation instead of individually. Since some participants were observing others performing the experiments, while the first participant in each batch performed the experiment without any prior experiences. This may bring some dependence among the participants and an order effect.

Fourth, the learning subject of the experiment was limited to the context of pedestrian road safety. Other learning subjects other than pedestrian safety may achieve different learning outcome when using VR as the learning tool. Also, the participants were invited to evaluate the learning system with a learning subject they already mastered (all college students know how to cross the road safely). The learning outcome from a learning subject completely novel to the participants may also be different.

Fifth, out of consideration of increasing the statistical power, this research used the fractional factorial design that only included the main effects and two-degree interactions in the experiment. Although the higher level of interactions will mostly make little differences, there is a risk that some significant higher order interactions may be missed.

6.4 Future Research

This study provides some issues that are worth further research.

First, the experiment can be replicated in another sample population other than the college students to evaluate if the conclusions drawn from this research can be generalized.

Second, the WIPLS system can be upgraded to include more customizability so that it can become capable of exploring more VR factors. This research used a flat screen for display and Kinect sensor as one of the control mechanisms. In the future, more customizable features can be added to the WIPLS, like the HMD as the display device, the Nintendo Wii handler as the controlling option, or the gaming steering wheel as the controller for driving related training, etc.

With the upgraded WIPLS, more comprehensive experiments involving more VR factors can be conducted, and correspondingly more insights on these VR factors can also be gained.

Third, the WIPLS can be modified to test a different learning subject. Since the software of WIPLS is developed using the Unity3D game engine which can build the virtual environment and the gaming logic with relatively short development cycle, more learning contexts can be added into the system.

Fourth, this research used fractional factorial design out of concerned for the statistic power being diminished when divided into too many treatment combinations. In the future, if a much bigger sample population can be recruited, it is recommended that a full factorial design can be used to test the higher degree of interactions between the VR factors, while at the same time maintaining a high statistic power.

6.5 Summary

This research extended the knowledge of using VR as a learning tool and further explored how the learning outcome was affected by the VR factors. A customizable WIPLS system, a survey instrument, and a theoretical model were developed to answer the research question: how does VR affect the learning outcome? Statistic model and hypotheses were formulated to explore the relationship between all these variables of interest, and an empirical experiment was conducted to collect the data on how participants were rating different sub-VR systems. Experiment results were analyzed from the collected data, and meaningful conclusions as well as insights were obtained and interpreted.

This study reported that natural semantics was the most significant VR factor that affected the perceived learning effectiveness of the educational VR system and the satisfaction towards the participants. Insignificant VR factors found in this research also provided insightful suggestions in guiding the development of educational VR applications. What's more, the hypotheses test on the correlations between the VR factors and the critical components of the theoretical learning frameworks supported the proposed research model. The findings from this research can provide insights and suggestions for the VR practitioners for their future VR design and development.

References

Adamovich, S. V., Fluet, G. G., Tunik, E., and Merians, A. S. (2009). Sensorimotor training in virtual reality: a review. *NeuroRehabilitation*, 25(1):29–44. 7, 8

Akaike, H. (1998). Information Theory and an Extension of the Maximum Likelihood Principle. In Parzen, E., Tanabe, K., and Kitagawa, G., editors, *Selected Papers of Hirotugu Akaike*, Springer Series in Statistics, pages 199–213. Springer New York. DOI: 10.1007/978-1-4612-1694-0_15. 66

Aljohani, N., Loke, S., and Ng, W. (2010). A Mobile-Based Group Quiz System to Promote Collaborative Learning and Facilitate Instant Feedback. 33

Alzantot, M. and Youssef, M. (2012). UPTIME: Ubiquitous pedestrian tracking using mobile phones. In *2012 IEEE Wireless Communications and Networking Conference (WCNC)*, pages 3204–3209. 40

Anderson, D. R. and Burnham, K. P. (2002). Avoiding Pitfalls When Using Information-Theoretic Methods. *The Journal of Wildlife Management*, 66(3):912–918. 66

Anderson, M. J. and Whitcomb, P. J. (2016). *DOE simplified: practical tools for effective experimentation*. CRC Press. 46

Bangay, S. and Preston, L. (1998). An investigation into factors influencing immersion in interactive virtual reality environments. *Studies in health technology and informatics*, pages 43–51. 9, 21, 99

Bearse, P. M. and Bozdogan, H. (1998). Subset selection in vector autoregressive models using the genetic algorithm with informational complexity as the fitness function. *Systems Analysis Modelling Simulation*, 31(1-2):61–91. 2. 66

Beauregard, S. (2006). A Helmet-Mounted Pedestrian Dead Reckoning System. In *2006 3rd International Forum on Applied Wearable Computing (IFAWC)*, pages 1–11. zcr-pdr. 40

BENBUNAN-FICH, R. and HILTZ, S. R. (2003). Mediators of the effectiveness of online courses. *IEEE Transactions on Professional Communication*, 46(4):298–312. 11

Blackledge, J. and Barrett, M. (2012). Evaluation of a Prototype Desktop Virtual Reality Model Developed to Enhance Electrical Safety and Design in the Built Environment. *Articles*. lit with factors. 22

Borer, S. (2010). A model of jerkiness for temporal impairments in video transmission. In *2010 Second International Workshop on Quality of Multimedia Experience (QoMEX)*, pages 218–223. 10

Bozdogan, H. (1987a). ICOMP: A new model-selection criterion. In *1. Conference of the International Federation of Classification Societies*, pages 599–608. 4. 66

Bozdogan, H. (1987b). Model selection and Akaike's information criterion (AIC): The general theory and its analytical extensions. *Psychometrika*, 52(3):345–370. 3. 66

Bozdogan, H. (1988). The theory and applications of information-theoretic measure of complexity (ICOMP) as a new model selection criterion. *Unpublished research report, the Institute of Statistical mathematics, Tokyo, Japan, and the Department of mathematics, University of Virginia, Charlottesville, VA*. 5. 66

Bozdogan, H. (1990). On the information-based measure of covariance complexity and its application to the evaluation of multivariate linear models. *Communications in Statistics-Theory and Methods*, 19(1):221–278. 6. 66

Bozdogan, H. and Haughton, D. M. A. (1998). Informational complexity criteria for regression models. *Computational Statistics & Data Analysis*, 28(1):51–76. 1. 66

Bricken, M. (1991). Virtual reality learning environments: potentials and challenges. *ACM SIGGRAPH Computer Graphics*, 25(3):178–184. 15

Brooks, J. G. and Brooks, M. G. (1999). *In search of understanding: The case for constructivist classrooms*. ASCD. x, 11, 12

Bruno, L., Pereira, J., and Jorge, J. (2013). A new approach to walking in place. In *Human-Computer InteractionCINTERACT 2013*, pages 370–387. Springer. 10

Burigat, S. and Chittaro, L. (2007). Navigation in 3d virtual environments: Effects of user experience and location-pointing navigation aids. *International Journal of Human-Computer Studies*, 65(11):945–958. 99

Cai, Y., Snel, I., Bharathi, B. S., Klein, C., and Klein-Seetharaman, J. (2003). Towards biomedical problem solving in a game environment. In *Computational SciencełICCS 2003*, pages 1005–1014. Springer. 3, 19

Charsky, D. (2010). From Edutainment to Serious Games: A Change in the Use of Game Characteristics. *Games and Culture*. 15

Chen, C. H. (1988). *Signal Processing Handbook*. CRC Press. 41

Chen, W., Chen, R., Chen, Y., Kuusniemi, H., and Wang, J. (2010). An effective Pedestrian Dead Reckoning algorithm using a unified heading error model. In *Position Location and Navigation Symposium (PLANS), 2010 IEEE/ION*, pages 340–347. zcr-pdr. 40

Choi, S. H. and Cheung, H. H. (2008). A versatile virtual prototyping system for rapid product development. *Computers in Industry*, 59(5):477–488. 14

Chou, S.-W. and Liu, C.-H. (2005). Learning effectiveness in a Web-based virtual learning environment: a learner control perspective. *Journal of computer assisted learning*, 21(1):65–76. 71

Coles, C. D., Strickland, D. C., Padgett, L., and Bellmoff, L. (2007). Games that work: Using computer games to teach alcohol-affected children about fire and street safety. *Research in Developmental Disabilities*, 28(5):518–530. 3, 23, 108

Coller, B. D. and Scott, M. J. (2009). Effectiveness of using a video game to teach a course in mechanical engineering. *Computers & Education*, 53(3):900–912. 14

Coon, D. and Mitterer, J. (2012). *Introduction to Psychology: Gateways to Mind and Behavior with Concept Maps and Reviews*. Cengage Learning. no full text. 13

Cronbach, L. J. (1957). The two disciplines of scientific psychology. *American psychologist*, 12(11):671. 54

Duh, H. B.-L., Lin, J. J., Kenyon, R. V., Parker, D. E., and Furness, T. A. (2001). Effects of field of view on balance in an immersive environment. In *Virtual Reality, 2001. Proceedings. IEEE*, pages 235–240. IEEE. 82

Ebner, M. and Holzinger, A. (2007). Successful implementation of user-centered game based learning in higher education: An example from civil engineering. *Computers & education*, 49(3):873–890. 3

Erhel, S. and Jamet, E. (2013). Digital game-based learning: Impact of instructions and feedback on motivation and learning effectiveness. *Computers & Education*, 67:156–167. 19

Feasel, J., Whitton, M. C., and Wendt, J. D. (2008). LLCM-WIP: Low-latency, continuous-motion walking-in-place. pages 97–104. IEEE. 10, 41

Felix, U. (2002). The web as a vehicle for constructivist approaches in language teaching. *ReCALL*, 14(01):2–15. 11, 13

Field, A. (2009). *Discovering statistics using SPSS*. Sage publications. 60

Glasersfeld, E. v. (1989). Cognition, construction of knowledge, and teaching. *Synthese*, 80(1):121–140. 1

Gutierrez, M., Vexo, F., and Thalmann, D. (2008). *Stepping into virtual reality*. Springer. 7

Holzinger, A., Pichler, A., and Maurer, H. (2006). Multi Media e-learning software TRIANGLE case-study: experimental results and lessons learned. *Journal of Universal Science and Technology of Learning*, 1(1):61–92. 15, 21

Hongbiao Yang, Shuguang Ji, E. R. W. R. S. (2015). Development of Walking in Place System based on Zero Crossing Algorithm. *The International Journal of Virtual Reality*, 15(02):30–43. 104

Huang, H.-M. (2002). Toward constructivism for adult learners in online learning environments. *British Journal of Educational Technology*, 33(1):27–37. 32

Huang, Y.-C., Backman, S. J., Chang, L.-L., Backman, K. F., and McGuire, F. A. (2013). Experiencing student learning and tourism training in a 3d virtual world: An exploratory study. *Journal of Hospitality, Leisure, Sport & Tourism Education*, 13:190–201. 14, 35

Huynh-Thu, Q. and Ghanbari, M. (2006). Impact of jitter and jerkiness on perceived video quality. 10

Istance, H., Hyrskykari, A., Vickers, S., and Chaves, T. (2009). For Your Eyes Only: Controlling 3d Online Games by Eye-Gaze. In Gross, T., Gulliksen, J., Kotz, P., Oestreicher, L., Palanque, P., Prates, R. O., and Winckler, M., editors, *Human-Computer Interaction C INTERACT 2009*, number 5726 in Lecture Notes in Computer Science, pages 314–327. Springer Berlin Heidelberg. 41

Jayaram, S., Connacher, H. I., and Lyons, K. W. (1997). Virtual assembly using virtual reality techniques. *Computer-Aided Design*, 29(8):575–584. 14

Johnson, D. W., Johnson, R. T., and Smith, K. A. (1998). Cooperative learning returns to college what evidence is there that it works? *Change: the magazine of higher learning*, 30(4):26–35. 10

Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement*, 20:141–151. 75

Kim, J., Gracanin, D., and Quek, F. (2012). Sensor-fusion walking-in-place interaction technique using mobile devices. pages 39–42. IEEE. 10

Kim, J. H., Jang, S. H., Kim, C. S., Jung, J. H., and You, J. H. (2009). Use of virtual reality to enhance balance and ambulation in chronic stroke: a double-blind, randomized controlled study. *American Journal of Physical Medicine & Rehabilitation*, 88(9):693–701. 14

Korolov, M. (2016). *Fiit VR has a wider field of view than many competitors, including the Samsung Gear VR.* xv, 101

Kuze, J. and Ukai, K. (2008). Subjective evaluation of visual fatigue caused by motion images. *Displays*, 29(2):159–166. 70

Lainema, T. and Makkonen, P. (2003). Applying constructivist approach to educational business games: Case REALGAME. *Simulation & gaming*, 34(1):131–149. 11

Lee, E. A. L. (2011). *An investigation into the effectiveness of virtual reality-based learning*. phd, Murdoch University. 8, 11, 71

Lefoe, G. (1998). Creating constructivist learning environments on the web: The challenge in higher education. In *Ascilite*, volume 98, page 453. 11

Lepper, M. R., Greene, D., and Nisbett, R. E. (1973). Undermining children's intrinsic interest with extrinsic reward: A test of the "overjustification" hypothesis. *Journal of Personality and Social Psychology*, 28(1):129–137. 13

Lin, W. and Jay Kuo, C. C. (2011). Perceptual visual quality metrics: A survey. *Journal of Visual Communication and Image Representation*, 22(4):297–312. 10

Maki, P. L. (2012). *Assessing for learning: Building a sustainable commitment across the institution*. Stylus Publishing, LLC. 10

Mallows, C. L. (1973). Some Comments on C p. *Technometrics*, 15(4):661–675. 66

Malone, T. W. and Lepper, M. R. (1987). Making learning fun: A taxonomy of intrinsic motivations for learning. *Aptitude, learning, and instruction*, 3(1987):223–253. 14

Mann, B. D., Eidelson, B. M., Fukuchi, S. G., Nissman, S. A., Robertson, S., and Jardines, L. (2002). The development of an interactive game-based tool for learning surgical management algorithms via computer. *The American Journal of Surgery*, 183(3):305–308. 14, 108

Marks, R. B., Sibley, S. D., and Arbaugh, J. B. (2005). A Structural Equation Model of Predictors for Effective Online Learning. *Journal of Management Education*, 29(4):531–563. 11

Martin, C., Burkert, D., Choi, K., Wieczorek, N., McGregor, P., Herrmann, R., and Beling, P. (2012). A real-time ergonomic monitoring system using the Microsoft Kinect. In *2012 IEEE Systems and Information Design Symposium (SIEDS)*, pages 50–55. 70

Mayer, R. E. (2004). Should there be a three-strikes rule against pure discovery learning? *American Psychologist*, 59(1):14. 19

Mayer, R. E., Mautone, P., and Prothero, W. (2002). Pictorial aids for learning by doing in a multimedia geology simulation game. *Journal of Educational Psychology*, 94(1):171–185. 20

Meir, A., Parmet, Y., and Oron-Gilad, T. (2013). Towards understanding child-pedestrians hazard perception abilities in a mixed reality dynamic environment. *Transportation research part F: traffic psychology and behaviour*, 20:90–107. 23

Meluso, A., Zheng, M., Spires, H. A., and Lester, J. (2012). Enhancing 5th graders science content knowledge and self-efficacy through game-based learning. *Computers & Education*, 59(2):497–504. 19

Mirelman, A., Patritti, B. L., Bonato, P., and Deutsch, J. E. (2010). Effects of virtual reality training on gait biomechanics of individuals post-stroke. *Gait & posture*, 31(4):433–437. 14

Moreno, R. and Mayer, R. E. (2005). Role of Guidance, Reflection, and Interactivity in an Agent-Based Multimedia Game. *Journal of educational psychology*, 97(1):117. 20

Moss, J., Scisco, J., and Muth, E. (2008). Simulator Sickness during Head Mounted Display (HMD) of Real World Video Captured Scenes. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 52(19):1631–1634. 70

Mujber, T. S., Szecsi, T., and Hashmi, M. S. J. (2004). Virtual reality applications in manufacturing process simulation. *Journal of Materials Processing Technology*, 155C156:1834–1838. 14, 102

Multon, F. and Olivier, A.-H. (2013). Biomechanics of Walking in Real World: Naturalness we Wish to Reach in Virtual Reality. In Steinicke, F., Visell, Y., Campos, J., and Lcuyer, A., editors, *Human Walking in Virtual Environments*, pages 55–77. Springer New York. 10

Nalbant, G. and Bostan, B. (2006). Interaction in virtual reality. In *4th international symposium of interactive medical design (ISIMD)*. 31

Nilsson, N. C., Serafin, S., Laursen, M. H., Pedersen, K. S., Sikstrom, E., and Nordahl, R. (2013). Tapping-in-place: Increasing the naturalness of immersive walking-in-place locomotion through novel gestural input. pages 31–38. IEEE. 69

Olive, J., Thouvenin, I., Lemasson, G., and Sbaouni, M. (2006). Tire Manufacturing supported by virtual environment. In *Virtual Reality International Conference*, volume 6. 14, 102

Pan, Z., Cheok, A. D., Yang, H., Zhu, J., and Shi, J. (2006). Virtual reality and mixed reality for virtual learning environments. *Computers & Graphics*, 30(1):20–28. 14

Pausch, R., Proffitt, D., and Williams, G. (1997). Quantifying Immersion in Virtual Reality. In *Proceedings of the 24th Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH '97, pages 13–18, New York, NY, USA. ACM Press/Addison-Wesley Publishing Co. 21, 31, 99

Piekarski, W., Gunther, B., and Thomas, B. (1999). Integrating virtual and augmented realities in an outdoor application. In *2nd IEEE and ACM International Workshop on Augmented Reality, 1999. (IWAR '99) Proceedings*, pages 45–54. 14

Pintrich, P. R. and de Groot, E. V. (1990). Motivational and self-regulated learning components of classroom academic performance. *Journal of Educational Psychology*, 82(1):33–40. no full text. 13

Pitcairn, T. K. and Edlmann, T. (2000). Individual differences in road crossing ability in young children and adults. *British Journal of Psychology*, 91:391. 2

Prince, M. (2004). Does Active Learning Work? A Review of the Research. *Journal of Engineering Education*, 93(3):223–231. 10

Prothero, J. D. and Hoffman, H. G. (1995). Widening the field-of-view increases the sense of presence in immersive virtual environments. *Human Interface Technology Laboratory Technical Report TR-95*, 2. 82

Psotka, J. (1995). Immersive training systems: Virtual reality and education and training. *Instructional science*, 23(5-6):405–431. 10, 99

Rahim, A. and Eliana, E. (2013). *Exploring a desktop virtual reality application for education: the perspectives of spatial knowledge acquisition and information integration*. Thesis, Lincoln University. 99

Rizzo, A. and Kim, G. J. (2005). A SWOT Analysis of the Field of Virtual Reality Rehabilitation and Therapy. *Presence: Teleoperators & Virtual Environments*, 14(2):119–146. 102

Runyan, C. W., Casteel, C., Perkis, D., Black, C., Marshall, S. W., Johnson, R. M., Coyne-Beasley, T., Waller, A. E., and Viswanathan, S. (2005). Unintentional injuries in the home in the United States: Part I: Mortality. *American journal of preventive medicine*, 28(1):73–79. 2

Ryan, R. M. and Deci, E. L. (2000). Intrinsic and Extrinsic Motivations: Classic Definitions and New Directions. *Contemporary Educational Psychology*, 25(1):54–67. 13

Ryan, R. M., Rigby, C. S., and Przybylski, A. (2006). The Motivational Pull of Video Games: A Self-Determination Theory Approach. *Motivation and Emotion*, 30(4):344–360. 99

Satava, R. M. (1995). Medical applications of virtual reality. *Journal of Medical Systems*, 19(3):275–280. 14

Saxena, N., Kyaw, B. M., Vseteckova, J., Dev, P., Paul, P., Lim, K. T. K., Kononowicz, A., Masiello, I., Tudor Car, L., Nikolaou, C. K., Zary, N., and Car, J. (2016). Virtual reality environments for health professional education. In *Cochrane Database of Systematic Reviews*. John Wiley & Sons, Ltd. 103

Schalkoff, R. J. (1989). *Digital image processing and computer vision*. John Wiley & Sons Australia, Limited. 38

Schwarz, G. (1978). Estimating the Dimension of a Model. *The Annals of Statistics*, 6(2):461–464. 66

Schwebel, D. C., Gaines, J., and Severson, J. (2008). Validation of virtual reality as a tool to understand and prevent child pedestrian injury. *Accident Analysis & Prevention*, 40(4):1394–1400. 24

Shin, S., Park, C., Kim, J., Hong, H., and Lee, J. (2007). Adaptive Step Length Estimation Algorithm Using Low-Cost MEMS Inertial Sensors. In *IEEE Sensors Applications Symposium, 2007. SAS '07*, pages 1–5. 40

Sibert, L. E., Templeman, J. N., Page, R. C., Barron, J. T., and McCune, J. A. (2004). Initial Assessment of Human Performance Using the Gaiter Interaction Technique to Control Locomotion in Fully Immersive Virtual Environments. Technical report. 10

Simpson, G., Johnston, L., and Richardson, M. (2003). An investigation of road crossing in a virtual environment. *Accident Analysis & Prevention*, 35(5):787–796. 23

Springer, L., Stanne, M. E., and Donovan, S. S. (1999). Effects of small-group learning on undergraduates in science, mathematics, engineering, and technology: A meta-analysis. *Review of educational research*, 69(1):21–51. 10

Stakem, F., AlRegib, G., and Juang, B.-H. F. (2007). Towards Modeling Human Arm Movement in a CVE. In *Proceedings of the First International Conference on Immersive Telecommunications*, ImmersCom '07, pages 8:1–8:5, ICST, Brussels, Belgium, Belgium. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering). 10

Templeman, J., Denbrook, P., and Sibert, L. (1999). Virtual locomotion: Walking in place through virtual environments. *Presence*, 8(6):598–617. 9

Usoh, M., Arthur, K., Whitton, M. C., Bastos, R., Steed, A., Slater, M., and Brooks, Jr., F. P. (1999). Walking ? Walking-in-place ? Flying, in Virtual Environments. In *Proceedings of the 26th Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH '99, pages 359–364, New York, NY, USA. ACM Press/Addison-Wesley Publishing Co. r[10]. 10, 69

Vallerand, R. J., Pelletier, L. G., Blais, M. R., Briere, N. M., Senecal, C., and Vallieres, E. F. (1992). The Academic Motivation Scale: A Measure of Intrinsic, Extrinsic, and Amotivation in Education. *Educational and Psychological Measurement*, 52(4):1003–1017. 13

Vera, L., Herrera, G., and Vived, E. (2005). Virtual Reality School for Children with Learning Difficulties. In *Proceedings of the 2005 ACM SIGCHI International Conference on Advances in Computer Entertainment Technology*, ACE '05, pages 338–341, New York, NY, USA. ACM. 14, 108

Virvou, M. and Katsionis, G. (2008). On the usability and likeability of virtual reality games for education: The case of VR-ENGAGE. *Computers & Education*, 50(1):154–178. 22

Virvou, M., Katsionis, G., and Manos, K. (2005). Combining Software Games with Education: Evaluation of its Educational Effectiveness. *Educational Technology & Society*, 8(2):54–65. 15, 22

Virvou, M., Manos, C., Katsionis, G., and Tourtoglou, K. (2002). VR-ENGAGE: A virtual reality educational game that incorporates intelligence. In *Proceedings of IEEE international conference on advanced learning technologies*, pages 16–19. 15

Vogel, J. J., Greenwood-Ericksen, A., Cannon-Bowers, J., and Bowers, C. A. (2006). Using Virtual Reality with and without Gaming Attributes for Academic Achievement. *Journal of Research on Technology in Education*, 39(1):105–118. 108

Vora, J., Nair, S., Gramopadhye, A. K., Duchowski, A. T., Melloy, B. J., and Kanki, B. (2002). Using virtual reality technology for aircraft visual inspection training: presence and comparison studies. *Applied Ergonomics*, 33(6):559–570. 21, 99, 108

Wade, E. and Winstein, C. J. (2011). Virtual reality and robotics for stroke rehabilitation: where do we go from here? *Topics in stroke rehabilitation*, 18(6):685–700. 14

Wendt, J. D., Whitton, M. C., and Brooks Jr, F. P. (2010). Gud wip: Gait-understanding-driven walking-in-place. pages 51–58. IEEE. 10

Whitton, M. C. and Peck, T. C. (2013). Stepping-driven locomotion interfaces. In *Human Walking in Virtual Environments*, pages 241–262. Springer. 10

Williams, B., Bailey, S., Narasimham, G., Li, M., and Bodenheimer, B. (2011). Evaluation of walking in place on a wii balance board to explore a virtual environment. *ACM Transactions on Applied Perception (TAP)*, 8(3):19. 10

Winn, W., Human, Center, W., and University (1993). A Conceptual Basis for Educational Applications of Virtual Reality. natural semantics, symbol. 8, 13, 14, 15, 31, 93, 94

Witmer, B. G. and Singer, M. J. (1998). Measuring Presence in Virtual Environments: A Presence Questionnaire. *Presence: Teleoperators and Virtual Environments*, 7(3):225–240. 45

Wobbrock, J. O., Findlater, L., Gergle, D., and Higgins, J. J. (2011). The Aligned Rank Transform for Nonparametric Factorial Analyses Using Only Anova Procedures. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '11, pages 143–146, New York, NY, USA. ACM. x, 61, 62

Wrzesien, M. and Raya, M. A. (2010). Learning in serious virtual worlds: Evaluation of learning effectiveness and appeal to students in the E-Junior project. *Computers & Education*, 55(1):178–187. 3, 15, 29

Yan, L., Allison, R., and Rushton, S. (2004). New Simple Virtual Walking Method -Walking on the Spot. In *8th Annual Immersive Projection Technology (IPT) Symposium Electronic Proceedings*. 10

Zhang, D., Zhou, L., Briggs, R. O., and Nunamaker Jr, J. F. (2006). Instructional video in e-learning: Assessing the impact of interactive video on learning effectiveness. *Information & Management*, 43(1):15–27. 33

Zielinski, D. J., McMahan, R. P., and Brady, R. B. (2011). Shadow walking: An unencumbered locomotion technique for systems with under-floor projection. pages 167–170. IEEE. 10

Zyda, M. (2005). From visual simulation to virtual reality to games. *Computer*, 38(9):25–32. 102

Appendices

A Hypotheses

A.1 Alternative Hypotheses

▶ *Visualization:*

- Hypothesis 1a (H_{1a}): Participants in VR subgroup with high level of visualization will give different rating on active learning than the VR subgroup with low level of visualization.
- Hypothesis 1b (H_{1b}): Participants in VR subgroup with high level of visualization will give different rating on interactive learning than the VR subgroup with low level of visualization.
- Hypothesis 1c (H_{1c}): Participants in VR subgroup with high level of visualization will give different rating on authentic problem than the VR subgroup with low level of visualization.
- Hypothesis 1d (H_{1d}): Participants in VR subgroup with high level of visualization will give different rating on control than the VR subgroup with low level of visualization.
- Hypothesis 1e (H_{1e}): Participants in VR subgroup with high level of visualization will give different rating on challenge than the VR subgroup with low level of visualization.
- Hypothesis 1f (H_{1f}): Participants in VR subgroup with high level of visualization will give different rating on experience than the VR subgroup with low level of visualization.

▶ *Natural Semantics:*

- Hypothesis 1g (H_{1g}): Participants in VR subgroup with high level of natural semantics will give different rating on active learning than the VR subgroup with low level of natural semantics.
- Hypothesis 1h (H_{1h}): Participants in VR subgroup with high level of natural semantics will give different rating on interactive learning than the VR subgroup with low level of natural semantics.
- Hypothesis 1i (H_{1i}): Participants in VR subgroup with high level of natural semantics will give different rating on authentic problem than the VR subgroup with low level of natural semantics.
- Hypothesis 1j (H_{1j}): Participants in VR subgroup with high level of natural semantics will give different rating on control than the VR subgroup with low level of natural semantics.

- Hypothesis 1k (H_{1k}): Participants in VR subgroup with high level of natural semantics will give different rating on challenge than the VR subgroup with low level of natural semantics.
- Hypothesis 1l (H_{1l}): Participants in VR subgroup with high level of natural semantics will give different rating on experience than the VR subgroup with low level of natural semantics.

▶ *Interactoin:*

- Hypothesis 1m H_{1m}: Participants in VR subgroup with high level of interaction will give different rating on active learning than the VR subgroup with low level of interaction.
- Hypothesis 1m (H_{1m}): Participants in VR subgroup with high level of interaction will give different rating on interactive learning than the VR subgroup with low level of interaction.
- Hypothesis 1o (H_{1o}): Participants in VR subgroup with high level of interaction will give different rating on authentic problem than the VR subgroup with low level of interaction.
- Hypothesis 1p (H_{1p}): Participants in VR subgroup with high level of interaction will give different rating on control than the VR subgroup with low level of interaction.
- Hypothesis 1q (H_{1q}): Participants in VR subgroup with high level of interaction will give different rating on challenge than the VR subgroup with low level of interaction.
- Hypothesis 1r (H_{1r}): Participants in VR subgroup with high level of interaction will give different rating on experience than the VR subgroup with low level of interaction.

▶ *Immersion:*

- Hypothesis 1s (H_{1s}): Participants in VR subgroup with high level of immersion will give different rating on active learning than the VR subgroup with low level of immersion.
- Hypothesis 1t (H_{1t}): Participants in VR subgroup with high level of immersion will give different rating on interactive learning than the VR subgroup with low level of immersion.
- Hypothesis 1u (H_{1u}): Participants in VR subgroup with high level of immersion will give different rating on authentic problem than the VR subgroup with low level of immersion.
- Hypothesis 1v (H_{1v}): Participants in VR subgroup with high level of immersion will give different rating on control than the VR subgroup with low level of immersion.
- Hypothesis 1w (H_{1w}): Participants in VR subgroup with high level of immersion will give different rating on challenge than the VR subgroup with low level of immersion.

- Hypothesis 1x (H_{1x}): Participants in VR subgroup with high level of immersion will give different rating on experience than the VR subgroup with low level of immersion.

A.2 Null Hypotheses

▶ *Visualization:*

- Hypothesis 0a (H_{0a}): Participants in VR subgroups with high level and low level of visualization will give the same rating on active learning.
- Hypothesis 0b (H_{0b}): Participants in VR subgroups with high level and low level of visualization will give the same rating on interactive learning.
- Hypothesis 0c (H_{0c}): Participants in VR subgroups with high level and low level of visualization will give the same rating on authentic problem.
- Hypothesis 0d (H_{0d}): Participants in VR subgroups with high level and low level of visualization will give the same rating on control.
- Hypothesis 0e (H_{0e}): Participants in VR subgroups with high level and low level of visualization will give the same rating on challenge.
- Hypothesis 0f (H_{0f}): Participants in VR subgroups with high level and low level of visualization will give the same rating on experience.

▶ *Natural Semantics:*

- Hypothesis 0g (H_{0g}): Participants in VR subgroups with high level and low level of natural semantics will give the same rating on active learning.
- Hypothesis 0h (H_{0h}): Participants in VR subgroups with high level and low level of natural semantics will give the same rating on interactive learning.
- Hypothesis 0i (H_{0i}): Participants in VR subgroups with high level and low level of natural semantics will give the same rating on authentic problem.
- Hypothesis 0j (H_{0j}): Participants in VR subgroups with high level and low level of natural semantics will give the same rating on control.
- Hypothesis 0k (H_{0k}): Participants in VR subgroups with high level and low level of natural semantics will give the same rating on challenge.

- Hypothesis 0l (H_{0l}): Participants in VR subgroups with high level and low level of natural semantics will give the same rating on experience.

▶ *Interactoin:*

- Hypothesis 0m (H_{0m}): Participants in VR subgroups with high level and low level of interaction will give the same rating on active learning.
- Hypothesis 0n (H_{0n}): Participants in VR subgroups with high level and low level of interaction will give the same rating on interactive learning.
- Hypothesis 0o (H_{0o}): Participants in VR subgroups with high level and low level of interaction will give the same rating on authentic problem.
- Hypothesis 0p (H_{0p}): Participants in VR subgroups with high level and low level of interaction will give the same rating on control.
- Hypothesis 0q (H_{0q}): Participants in VR subgroups with high level and low level of interaction will give the same rating on challenge.
- Hypothesis 0r (H_{0r}): Participants in VR subgroups with high level and low level of interaction will give the same rating on experience.

▶ *Immersion:*

- Hypothesis 0s (H_{0s}): Participants in VR subgroups with high level and low level of immersion will give the same rating on active learning.
- Hypothesis 0t (H_{0t}): Participants in VR subgroups with high level and low level of immersion will give the same rating on interactive learning.
- Hypothesis 0u (H_{0u}): Participants in VR subgroups with high level and low level of immersion will give the same rating on authentic problem.
- Hypothesis 0v (H_{0v}): Participants in VR subgroups with high level and low level of immersion will give the same rating on control.
- Hypothesis 0w (H_{0w}): Participants in VR subgroups with high level and low level of immersion will give the same rating on challenge.
- Hypothesis 0x (H_{0x}): Participants in VR subgroups with high level and low level of immersion will give the same rating on experience.

B Survey Instrument

Virtual Reality for Learning

1. What your VR task number? *

   _____

2. What is your gender? *

   Male Female

3. What is your age? *

   _____

4. Roughly how many hours have you spent playing video games for every week on average (e.g. gaming consoles, mobile phones, computers, etc.)? *

   None

   1 to 3 hours

   4 to 6 hours

   7 to 9 hours

   10 hours or more

5. How much do you know about Virtual Reality? *

   Never heard of

   Know about it but never experienced one by myself

   Used VR a couple of times

   I am an expert on this topic

6. I was more interested to learn the topics *

   Choose from 1-7, 7 means strongly agree, 1 means strongly disagree

7. I learned a lot of factual information in the topics *

   Choose from 1-7, 7 means strongly agree, 1 means strongly disagree

8. I gained a good understanding of the basic concepts of the materials *

   Choose from 1-7, 7 means strongly agree, 1 means strongly disagree

9. I learned to identify the main and important issues of the topics *

   Choose from 1-7, 7 means strongly agree, 1 means strongly disagree

10. I was interested and stimulated to learn more *

    Choose from 1-7, 7 means strongly agree, 1 means strongly disagree

11. I was able to summarize and concluded what I learned *

    Choose from 1-7, 7 means strongly agree, 1 means strongly disagree

12. The learning activities were meaningful *

    Choose from 1-7, 7 means strongly agree, 1 means strongly disagree

13. What I learned, I can apply in real context *

    Choose from 1-7, 7 means strongly agree, 1 means strongly disagree

14. The learning experience with the VR learning environment was better than that with the traditional classroom *

    Choose from 1-7, 7 means strongly agree, 1 means strongly disagree

15. I think this type of VR learning environment would benefit me for my learning achievement *

    Choose from 1-7, 7 means strongly agree, 1 means strongly disagree

16. I was satisfied with this type of VR learning *

    Choose from 1-7, 7 means strongly agree, 1 means strongly disagree

17. I was satisfied with the overall learning effectiveness *

    Choose from 1-7, 7 means strongly agree, 1 means strongly disagree

18. This VR system engages me to learn more proactively *

    Choose from 1-7, 7 means strongly agree, 1 means strongly disagree

19. I can interact with the VR system freely *

    Choose from 1-7, 7 means strongly agree, 1 means strongly disagree

20. The scenario and mechanism presented in this virtual world feels close to the real world *

    Choose from 1-7, 7 means strongly agree, 1 means strongly disagree

21. I have good control over this VR system *

    Choose from 1-7, 7 means strongly agree, 1 means strongly disagree

22. The task in this VR is challenging for me *

    Choose from 1-7, 7 means strongly agree, 1 means strongly disagree

23. The VR system provides great experience to me *

    Choose from 1-7, 7 means strongly agree, 1 means strongly disagree

C IRB Approval

C.1 UTK IRB Approval Letter

THE UNIVERSITY OF TENNESSEE KNOXVILLE

June 30, 2016

Re: UTK IRB-16-03058XP
Study Title: Exploring a semi virtual reality system for constructivist based education

Dear Hongbiao Yang:

The UTK Institutional Review Board (IRB) reviewed your application for the above referenced project. It determined that your application is eligible for expedited review under 45 CFR 46.110(b)(1), category (7). The IRB has reviewed these materials and determined that they do comply with proper consideration for the rights and welfare of human subjects and the regulatory requirements for the protection of human subjects. Therefore, this letter constitutes full approval by the IRB of your application (version 1.4) as submitted, including Informed Consent (v1.4), Questionnaires (v4.0) and the Recruitment Letter to Participants (v2.0). The forms have been dated and stamped IRB approved. Approval of this study will be valid from June 30, 2016 to June 29, 2017.

In the event that subjects are to be recruited using solicitation materials, such as brochures, posters, web-based advertisements, etc., these materials must receive prior approval of the IRB. Any revisions in the approved application must also be submitted to and approved by the IRB prior to implementation. In addition, you are responsible for reporting any unanticipated serious adverse events or other problems involving risks to subjects or others in the manner required by the local IRB policy.

Finally, re-approval of your project is required by the IRB in accord with the conditions specified above. You may not continue the research study beyond the time or other limits specified unless you obtain prior written approval of the IRB.

Sincerely,

Colleen P. Gilrane

Colleen P. Gilrane, Ph.D.
Chair

Institutional Review Board | Office of Research & Engagement
1534 White Avenue Knoxville, TN 37996-1529
865-974-7697 865-974-7400 fax irb.utk.edu

BIG ORANGE. BIG IDEAS.
Flagship Campus of the University of Tennessee System

A very small percentage of people have experienced photosensitive epileptic seizures when exposed to certain visual images, including flashing lights or patterns that may be appear in video games or other video displays. If you have a cardiac pacemaker or similar device, you should maintain a distance of 6 inches from the equipment due to potential interference from the equipment's radio-frequency emissions.

The VR program will be displayed on a flat TV screen, the motion sickness and discomfort caused by the VR program should be minimal. If you feel discomfort at any time of the experiment, please stop the participation and notify the instructor immediately.

PARTICIPATION AND BENEFITS:

Your participation is voluntary and you can decline to participate or end your participation at any time with no penalty or loss of benefits to which you are otherwise entitled. There are no direct benefits to you. Your participation in this research may benefit the study of exploring which factors are important in VR learning and how those factors are impacting the learning effectiveness.

CONTACT INFORMATION:

If you have questions at any time about the study or experience any problems, you may contact the researcher, (Hongbiao Yang), at Address: 865 Neyland Dr, Knoxville, TN 37996, and Office Phone Number at (865)-974-7655, or contact Dr Rupy Sawhney at Address: 865 Neyland Dr, Knoxville, TN 37996, and Office Phone Number at (865)-974-7653. If you have questions about your rights as a participant, contact the University of Tennessee, Knoxville IRB Compliance Officer at (865) 974-7697, email: utkirb@utk.edu.

CONSENT:

I have read the above information. I have received a copy of this form. I agree to participate in this study.

Participant's signature ______________________________ Date ______________________

IRB NUMBER: UTK IRB-16-03058XP
IRB APPROVAL DATE: 06/30/2016
IRB EXPIRATION DATE: 06/29/2017

C.2 Recruitment Letter to Participants

Dear Engineering Students,

I am a doctoral student of Department of Industrial and Systems Engineering at the University of Tennessee, Knoxville. I am conducting a study to explore how each factor of the Virtual Reality can impact the learning effectiveness among college students for my dissertation and ask for your participation. I will be working with Dr. Rupy Sawhney to implement my study. Virtual Reality for Education is a Virtual Reality Learning Environment system that is used to provide a virtual environment for the users and enable them to learn new knowledge and skills through a simulated scenario that is similar to real world environment. The whole participation (20 minutes) will include one practice session, one experiment session and one survey. A VR system named Walk in Place Learning System (WIPLS) will be used in this study. Participation is voluntary and you have the opinion to end your participation at any time. In the first session (2 minutes), you will practice the WIPLS. The purpose of this phase of experiment is to help you get familiar with the WIPLS. The instructor will briefly explain how to operate the WIPLS. If you have any questions during this practice session, please feel free to ask the instructor for help. They will have a 5 minutes break after the first session is over. In the second session (3 minutes), you will then be assigned to a particular task of experiments. Each task will be similar to but not exactly the same as the practice session. In this experiment session, each of you will use either the body movement or a controller app on a smartphone to control a virtual character to cross the road in a virtual environment. You should make your judgement and take your action accordingly in order to arrive at the other side of road safely. After you complete those two sessions, you will complete a survey (10 minutes) regarding your opinion about the experiment session as well as the learning effectiveness you perceive. The survey will ask demographic questions, but no identifiable information will be asked for. If you would like to participate in this study, please contact Hongbiao Yang (865-246-8741) or Rupy Sawhney (865-974-7653) for more information. Agreement of your participation implies your consent. If you have questions about the experiment at any time, please feel free to contact Hongbiao Yang (hyang22@vols.utk.edu). If you have questions about your rights as a participant, contact the University of Tennessee, Knoxville IRB Compliance Officer at (865) 974-7697, email:

utkirb@utk.edu. Thank you in advance for your participation. Your help is greatly appreciated and critical to this study!

Sincerely,

Hongbiao Yang

The University of Tennessee

Department of Industrial & Systems Engineering

D ANOVA Table on Variables after ART

D.1 ANOVA Tables with Perceived Learning Effectiveness as the Response Variable

Table D.1: ANOVA table on variable X1Vi with perceived learning effectiveness

Dependent Variable: ARTLearnEffforX1Vi					
S Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	5616.394a	10	561.639	0.112	1
Intercept	3084653	1	3084653	616.197	0
X1Vi * X2NS	10	1	10	0.002	0.964
X1Vi * X3In	260.1	1	260.1	0.052	0.82
X1Vi * X4Im	123.019	1	123.019	0.025	0.876
X2NS * X3In	58.806	1	58.806	0.012	0.914
X2NS * X4Im	363.006	1	363.006	0.073	0.788
X3In * X4Im	684.756	1	684.756	0.137	0.712
X1Vi	**1836.025**	**1**	**1836.025**	**0.367**	**0.545**
X2NS	1066.056	1	1066.056	0.213	0.645
X3In	257.556	1	257.556	0.051	0.821
X4Im	200.256	1	200.256	0.04	0.842
Error	1146363.6	229	5005.955		
Total	4636840	240			
Corrected Total	1151980	239			
a. R Squared = .005 (Adjusted R Squared = -.039)					

Table D.2: ANOVA table on variable X2NS with perceived learning effectiveness

Dependent Variable: ARTLearnEffforX2NS					
S Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	21222.444a	10	2122.244	0.43	0.931
Intercept	3087218.8	1	3087218.8	625.221	0
X1Vi * X2NS	3.906	1	3.906	0.001	0.978
X1Vi * X3In	406.406	1	406.406	0.082	0.774
X1Vi * X4Im	79.219	1	79.219	0.016	0.899
X2NS * X3In	43.056	1	43.056	0.009	0.926
X2NS * X4Im	372.1	1	372.1	0.075	0.784
X3In * X4Im	585.225	1	585.225	0.119	0.731
X1Vi	47.306	1	47.306	0.01	0.922
X2NS	**16463.306**	**1**	**16463.306**	**3.334**	**0.069**
X3In	247.506	1	247.506	0.05	0.823
X4Im	198.025	1	198.025	0.04	0.841
Error	1130757.6	229	4937.806		
Total	4636840	240			
Corrected Total	1151980	239			
a. R Squared = .018 (Adjusted R Squared = -.024)					

Table D.3: ANOVA table on variable X1Vi*X2NS with perceived learning effectiveness

Dependent Variable: ARTLearnEffforX1ViX2NS					
S Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	28261.844a	10	2826.184	0.576	0.833
Intercept	3086898	1	3086898	629.072	0
X1Vi * X2NS	**21045.156**	**1**	**21045.156**	**4.289**	**0.039**
X1Vi * X3In	278.256	1	278.256	0.057	0.812
X1Vi * X4Im	84.169	1	84.169	0.017	0.896
X2NS * X3In	3.906	1	3.906	0.001	0.978
X2NS * X4Im	469.225	1	469.225	0.096	0.757
X3In * X4Im	950.625	1	950.625	0.194	0.66
X1Vi	5.256	1	5.256	0.001	0.974
X2NS	1339.806	1	1339.806	0.273	0.602
X3In	636.006	1	636.006	0.13	0.719
X4Im	403.225	1	403.225	0.082	0.775
Error	1123718.2	229	4907.066		
Total	4636840	240			
Corrected Total	1151980	239			
a. R Squared = .025 (Adjusted R Squared = -.018)					

Table D.4: ANOVA table on variable X3In with perceived learning effectiveness

Dependent Variable: ARTLearnEffforX3In					
S Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	15545.000a	10	1554.5	0.313	0.977
Intercept	3087379.2	1	3087379.2	622.13	0
X1Vi * X2NS	3.025	1	3.025	0.001	0.98
X1Vi * X3In	235.225	1	235.225	0.047	0.828
X1Vi * X4Im	76.8	1	76.8	0.015	0.901
X2NS * X3In	13.225	1	13.225	0.003	0.959
X2NS * X4Im	532.9	1	532.9	0.107	0.743
X3In * X4Im	837.225	1	837.225	0.169	0.682
X1Vi	34.225	1	34.225	0.007	0.934
X2NS	1102.5	1	1102.5	0.222	0.638
X3In	**7209.225**	**1**	**7209.225**	**1.453**	**0.229**
X4Im	291.6	1	291.6	0.059	0.809
Error	1136435	229	4962.598		
Total	4636840	240			
Corrected Total	1151980	239			
a. R Squared = .013 (Adjusted R Squared = -.030)					

Table D.5: ANOVA table on variable X1Vi*X3In with perceived learning effectiveness

Dependent Variable: ARTLearnEffforX1ViX3In					
S Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	6151.844a	10	615.184	0.123	1
Intercept	3086898	1	3086898	616.933	0
X1Vi * X2NS	0.156	1	0.156	0	0.996
X1Vi * X3In	**1161.006**	**1**	**1161.006**	**0.232**	**0.63**
X1Vi * X4Im	84.169	1	84.169	0.017	0.897
X2NS * X3In	37.056	1	37.056	0.007	0.931
X2NS * X4Im	518.4	1	518.4	0.104	0.748
X3In * X4Im	577.6	1	577.6	0.115	0.734
X1Vi	20.306	1	20.306	0.004	0.949
X2NS	752.556	1	752.556	0.15	0.699
X3In	310.806	1	310.806	0.062	0.803
X4Im	148.225	1	148.225	0.03	0.863
Error	1145828.2	229	5003.616		
Total	4636840	240			
Corrected Total	1151980	239			
a. R Squared = .005 (Adjusted R Squared = -.038)					

Table D.6: ANOVA table on variable X2NS*X3In with perceived learning effectiveness

Dependent Variable: ARTLearnEffforX2NSX3In					
S Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	14536.000a	10	1453.6	0.293	0.982
Intercept	3087379.2	1	3087379.2	621.578	0
X1Vi * X2NS	1.225	1	1.225	0	0.987
X1Vi * X3In	308.025	1	308.025	0.062	0.804
X1Vi * X4Im	76.8	1	76.8	0.015	0.901
X2NS * X3In	**8880.4**	**1**	**8880.4**	**1.788**	**0.183**
X2NS * X4Im	390.625	1	390.625	0.079	0.779
X3In * X4Im	731.025	1	731.025	0.147	0.702
X1Vi	4.225	1	4.225	0.001	0.977
X2NS	1000	1	1000	0.201	0.654
X3In	360	1	360	0.072	0.788
X4Im	216.225	1	216.225	0.044	0.835
Error	1137444	229	4967.004		
Total	4636840	240			
Corrected Total	1151980	239			
a. R Squared = .013 (Adjusted R Squared = -.030)					

Table D.7: ANOVA table on variable X4Im with perceived learning effectiveness

Dependent Variable: ARTLearnEffforX4Im					
S Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	2304.844a	10	230.484	0.046	1
Intercept	3086256.5	1	3086256.5	614.741	0
X1Vi * X2NS	19.6	1	19.6	0.004	0.95
X1Vi * X3In	207.025	1	207.025	0.041	0.839
X1Vi * X4Im	94.519	1	94.519	0.019	0.891
X2NS * X3In	23.256	1	23.256	0.005	0.946
X2NS * X4Im	387.506	1	387.506	0.077	0.781
X3In * X4Im	787.656	1	787.656	0.157	0.692
X1Vi	13.225	1	13.225	0.003	0.959
X2NS	1076.406	1	1076.406	0.214	0.644
X3In	288.906	1	288.906	0.058	0.811
X4Im	**1.806**	**1**	**1.806**	**0**	**0.985**
Error	1149675.2	229	5020.416		
Total	4636840	240			
Corrected Total	1151980	239			
a. R Squared = .002 (Adjusted R Squared = -.042)					

Table D.8: ANOVA table on variable X1Vi*X4Im with perceived learning effectiveness

Dependent Variable: ARTLearnEffforX1ViX4Im					
S Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	5249.094a	10	524.909	0.105	1
Intercept	3031653.4	1	3031653.4	605.415	0
X1Vi * X2NS	28.056	1	28.056	0.006	0.94
X1Vi * X3In	223.256	1	223.256	0.045	0.833
X1Vi * X4Im	**3198.169**	**1**	**3198.169**	**0.639**	**0.425**
X2NS * X3In	15.006	1	15.006	0.003	0.956
X2NS * X4Im	302.5	1	302.5	0.06	0.806
X3In * X4Im	632.025	1	632.025	0.126	0.723
X1Vi	213.906	1	213.906	0.043	0.836
X2NS	945.756	1	945.756	0.189	0.664
X3In	182.756	1	182.756	0.036	0.849
X4Im	3.025	1	3.025	0.001	0.98
Error	1146730.9	229	5007.559		
Total	4636840	240			
Corrected Total	1151980	239			
a. R Squared = .005 (Adjusted R Squared = -.039)					

Table D.9: ANOVA table on variable X2NS*X4IM with perceived learning effectiveness

Dependent Variable: ARTLearnEffforX2NSX4Im					
S Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	3570.075a	10	357.008	0.071	1
Intercept	3085454.7	1	3085454.7	615.259	0
X1Vi * X2NS	4.556	1	4.556	0.001	0.976
X1Vi * X3In	242.556	1	242.556	0.048	0.826
X1Vi * X4Im	108.3	1	108.3	0.022	0.883
X2NS * X3In	7.225	1	7.225	0.001	0.97
X2NS * X4Im	**1204.506**	**1**	**1204.506**	**0.24**	**0.625**
X3In * X4Im	761.256	1	761.256	0.152	0.697
X1Vi	0.156	1	0.156	0	0.996
X2NS	1050.625	1	1050.625	0.21	0.648
X3In	280.9	1	280.9	0.056	0.813
X4Im	257.556	1	257.556	0.051	0.821
Error	1148409.9	229	5014.891		
Total	4636840	240			
Corrected Total	1151980	239			
a. R Squared = .003 (Adjusted R Squared = -.040)					

Table D.10: ANOVA table on variable X3In*X4Im with perceived learning effectiveness

Dependent Variable: ARTLearnEffforX3InX4Im					
S Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	2904.944a	10	290.494	0.058	1
Intercept	3084973.7	1	3084973.7	614.807	0
X1Vi * X2NS	1.225	1	1.225	0	0.988
X1Vi * X3In	189.225	1	189.225	0.038	0.846
X1Vi * X4Im	117.019	1	117.019	0.023	0.879
X2NS * X3In	31.506	1	31.506	0.006	0.937
X2NS * X4Im	486.506	1	486.506	0.097	0.756
X3In * X4Im	**15.006**	**1**	**15.006**	**0.003**	**0.956**
X1Vi	12.1	1	12.1	0.002	0.961
X2NS	1128.906	1	1128.906	0.225	0.636
X3In	333.506	1	333.506	0.066	0.797
X4Im	247.506	1	247.506	0.049	0.824
Error	1149075.1	229	5017.795		
Total	4636840	240			
Corrected Total	1151980	239			
a. R Squared = .003 (Adjusted R Squared = -.041)					

D.2 ANOVA Tables with Satisfaction as the Response Variable

Table D.11: ANOVA table on variable X1Vi with satisfaction

Dependent Variable: ARTSastisfactoryforX1Vi					
S Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	5567.000a	10	556.7	0.111	1
Intercept	3080646.1	1	3080646.1	615.37	0
X1Vi	**3413.256**	**1**	**3413.256**	**0.682**	**0.41**
X2NS	202.5	1	202.5	0.04	0.841
X3In	722.5	1	722.5	0.144	0.704
X4Im	620.156	1	620.156	0.124	0.725
X1Vi * X2NS	43.056	1	43.056	0.009	0.926
X1Vi * X3In	0.056	1	0.056	0	0.997
X1Vi * X4Im	210.675	1	210.675	0.042	0.838
X2NS * X3In	555.025	1	555.025	0.111	0.739
X2NS * X4Im	381.306	1	381.306	0.076	0.783
X3In * X4Im	195.806	1	195.806	0.039	0.843
Error	1146413	229	5006.17		
Total	4636840	240			
Corrected Total	1151980	239			
a. R Squared = .005 (Adjusted R Squared = -.039)					

Table D.12: ANOVA table on variable X2NS with satisfaction

Dependent Variable: ARTSastisfactoryforX2NS					
S Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	37508.544a	10	3750.854	0.771	0.657
Intercept	3081447.3	1	3081447.3	633.171	0
X1Vi	726.756	1	726.756	0.149	0.7
X2NS	**22681.406**	**1**	**22681.406**	**4.661**	**0.032**
X3In	832.656	1	832.656	0.171	0.68
X4Im	1040.4	1	1040.4	0.214	0.644
X1Vi * X2NS	79.806	1	79.806	0.016	0.898
X1Vi * X3In	43.056	1	43.056	0.009	0.925
X1Vi * X4Im	191.269	1	191.269	0.039	0.843
X2NS * X3In	1271.256	1	1271.256	0.261	0.61
X2NS * X4Im	160	1	160	0.033	0.856
X3In * X4Im	50.625	1	50.625	0.01	0.919
Error	1114471.5	229	4866.688		
Total	4636840	240			
Corrected Total	1151980	239			
a. R Squared = .033 (Adjusted R Squared = -.010)					

Table D.13: ANOVA table on variable X1Vi*X2NS with satisfaction

Dependent Variable: ARTSastisfactoryforX1ViX2NS					
S Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	17779.994a	10	1777.999	0.359	0.963
Intercept	3073760.3	1	3073760.3	620.606	0
X1Vi	680.625	1	680.625	0.137	0.711
X2NS	479.556	1	479.556	0.097	0.756
X3In	761.256	1	761.256	0.154	0.695
X4Im	543.906	1	543.906	0.11	0.741
X1Vi * X2NS	**13727.025**	**1**	**13727.025**	**2.772**	**0.097**
X1Vi * X3In	62.5	1	62.5	0.013	0.911
X1Vi * X4Im	416.269	1	416.269	0.084	0.772
X2NS * X3In	529.256	1	529.256	0.107	0.744
X2NS * X4Im	566.256	1	566.256	0.114	0.736
X3In * X4Im	299.756	1	299.756	0.061	0.806
Error	1134200	229	4952.838		
Total	4636840	240			
Corrected Total	1151980	239			
a. R Squared = .015 (Adjusted R Squared = -.028)					

Table D.14: ANOVA table on variable X3In with satisfaction

Dependent Variable: ARTSastisfactoryforX3In					
S Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	10623.594a	10	1062.359	0.213	0.995
Intercept	3072160	1	3072160	616.393	0
X1Vi	888.306	1	888.306	0.178	0.673
X2NS	97.656	1	97.656	0.02	0.889
X3In	**4171.806**	**1**	**4171.806**	**0.837**	**0.361**
X4Im	592.9	1	592.9	0.119	0.73
X1Vi * X2NS	16.256	1	16.256	0.003	0.955
X1Vi * X3In	2.256	1	2.256	0	0.983
X1Vi * X4Im	474.019	1	474.019	0.095	0.758
X2NS * X3In	486.506	1	486.506	0.098	0.755
X2NS * X4Im	656.1	1	656.1	0.132	0.717
X3In * X4Im	207.025	1	207.025	0.042	0.839
Error	1141356.4	229	4984.089		
Total	4636840	240			
Corrected Total	1151980	239			
a. R Squared = .009 (Adjusted R Squared = -.034)					

Table D.15: ANOVA table on variable X1Vi*X3In with satisfaction

Dependent Variable: ARTSastisfactoryforX1ViX3In					
S Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	6357.344a	10	635.734	0.127	0.999
Intercept	3080165.4	1	3080165.4	615.698	0
X1Vi	739.6	1	739.6	0.148	0.701
X2NS	257.556	1	257.556	0.051	0.821
X3In	770.006	1	770.006	0.154	0.695
X4Im	726.756	1	726.756	0.145	0.703
X1Vi * X2NS	90	1	90	0.018	0.893
X1Vi * X3In	**1199.025**	**1**	**1199.025**	**0.24**	**0.625**
X1Vi * X4Im	222.769	1	222.769	0.045	0.833
X2NS * X3In	604.506	1	604.506	0.121	0.728
X2NS * X4Im	363.006	1	363.006	0.073	0.788
X3In * X4Im	218.556	1	218.556	0.044	0.835
Error	1145622.7	229	5002.719		
Total	4636840	240			
Corrected Total	1151980	239			
a. R Squared = .006 (Adjusted R Squared = -.038)					

Table D.16: ANOVA table on variable X2NS*X3In with satisfaction

Dependent Variable: ARTSastisfactoryforX2NSX3In					
S Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	6904.644a	10	690.464	0.138	0.999
Intercept	3075681.1	1	3075681.1	615.096	0
X1Vi	796.556	1	796.556	0.159	0.69
X2NS	158.006	1	158.006	0.032	0.859
X3In	660.156	1	660.156	0.132	0.717
X4Im	714.025	1	714.025	0.143	0.706
X1Vi * X2NS	74.256	1	74.256	0.015	0.903
X1Vi * X3In	2.756	1	2.756	0.001	0.981
X1Vi * X4Im	351.919	1	351.919	0.07	0.791
X2NS * X3In	**975.156**	**1**	**975.156**	**0.195**	**0.659**
X2NS * X4Im	348.1	1	348.1	0.07	0.792
X3In * X4Im	133.225	1	133.225	0.027	0.87
Error	1145075.4	229	5000.329		
Total	4636840	240			
Corrected Total	1151980	239			
a. R Squared = .006 (Adjusted R Squared = -.037)					

Table D.17: ANOVA table on variable X4Im with satisfaction

Dependent Variable: ARTSastisfactoryforX4Im					
S Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	3765.400a	10	376.54	0.075	1
Intercept	3080646.1	1	3080646.1	614.404	0
X1Vi	841.806	1	841.806	0.168	0.682
X2NS	245.025	1	245.025	0.049	0.825
X3In	680.625	1	680.625	0.136	0.713
X4Im	**1339.806**	**1**	**1339.806**	**0.267**	**0.606**
X1Vi * X2NS	58.806	1	58.806	0.012	0.914
X1Vi * X3In	5.256	1	5.256	0.001	0.974
X1Vi * X4Im	210.675	1	210.675	0.042	0.838
X2NS * X3In	403.225	1	403.225	0.08	0.777
X2NS * X4Im	345.156	1	345.156	0.069	0.793
X3In * X4Im	247.506	1	247.506	0.049	0.824
Error	1148214.6	229	5014.038		
Total	4636840	240			
Corrected Total	1151980	239			
a. R Squared = .003 (Adjusted R Squared = -.040)					

Table D.18: ANOVA table on variable X1Vi*X4Im with satisfaction

Dependent Variable: ARTSastisfactoryforX1ViX4Im					
S Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	5100.944a	10	510.094	0.102	1
Intercept	3057138	1	3057138	610.426	0
X1Vi	1494.506	1	1494.506	0.298	0.585
X2NS	200.256	1	200.256	0.04	0.842
X3In	693.056	1	693.056	0.138	0.71
X4Im	1357.225	1	1357.225	0.271	0.603
X1Vi * X2NS	49.506	1	49.506	0.01	0.921
X1Vi * X3In	6.006	1	6.006	0.001	0.972
X1Vi * X4Im	**1200.169**	**1**	**1200.169**	**0.24**	**0.625**
X2NS * X3In	452.256	1	452.256	0.09	0.764
X2NS * X4Im	360	1	360	0.072	0.789
X3In * X4Im	220.9	1	220.9	0.044	0.834
Error	1146879.1	229	5008.205		
Total	4636840	240			
Corrected Total	1151980	239			
a. R Squared = .004 (Adjusted R Squared = -.039)					

Table D.19: ANOVA table on variable X2NS*X4Im with satisfaction

Dependent Variable: ARTSastisfactoryforX2NSX4Im					
S Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	3713.994a	10	371.399	0.074	1
Intercept	3081767.8	1	3081767.8	614.6	0
X1Vi	902.5	1	902.5	0.18	0.672
X2NS	232.806	1	232.806	0.046	0.83
X3In	770.006	1	770.006	0.154	0.696
X4Im	581.406	1	581.406	0.116	0.734
X1Vi * X2NS	32.4	1	32.4	0.006	0.936
X1Vi * X3In	3.6	1	3.6	0.001	0.979
X1Vi * X4Im	183.769	1	183.769	0.037	0.848
X2NS * X3In	500.556	1	500.556	0.1	0.752
X2NS * X4Im	**851.006**	**1**	**851.006**	**0.17**	**0.681**
X3In * X4Im	200.256	1	200.256	0.04	0.842
Error	1148266	229	5014.262		
Total	4636840	240			
Corrected Total	1151980	239			
a. R Squared = .003 (Adjusted R Squared = -.040)					

Table D.20: ANOVA table on variable X3In*X4Im with satisfaction

Dependent Variable: ARTSastisfactoryforX3InX4Im					
Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	7622.494a	10	762.249	0.153	0.999
Intercept	3080165.4	1	3080165.4	616.379	0
X1Vi	897.756	1	897.756	0.18	0.672
X2NS	283.556	1	283.556	0.057	0.812
X3In	851.006	1	851.006	0.17	0.68
X4Im	275.625	1	275.625	0.055	0.815
X1Vi * X2NS	45.156	1	45.156	0.009	0.924
X1Vi * X3In	9.506	1	9.506	0.002	0.965
X1Vi * X4Im	222.769	1	222.769	0.045	0.833
X2NS * X3In	322.056	1	322.056	0.064	0.8
X2NS * X4Im	348.1	1	348.1	0.07	0.792
X3In * X4Im	**2449.225**	**1**	**2449.225**	**0.49**	**0.485**
Error	1144357.5	229	4997.194		
Total	4636840	240			
Corrected Total	1151980	239			
a. R Squared = .007 (Adjusted R Squared = -.037)					

D.3 Statistics of Mann-Whitney U Tests

Table D.21: Mann Whitney U test of Active Learning with Visualization as grouping variable (H_{1a})

	Active Learning
Mann-Whitney U	6430
Wilcoxon W	13690
Z	-1.481
Asymp. Sig. (2-tailed)	0.139

Table D.22: Mann Whitney U test of Interactive Learning with Visualization as grouping variable (H_{1b})

	Interactive Learning
Mann-Whitney U	7057.5
Wilcoxon W	14317.5
Z	-0.272
Asymp. Sig. (2-tailed)	0.785

Table D.23: Mann Whitney U test of Authentic Problem with Visualization as grouping variable (H_{1c})

	Authentic Problem
Mann-Whitney U	7064.5
Wilcoxon W	14324.5
Z	-0.259
Asymp. Sig. (2-tailed)	0.796

Table D.24: Mann Whitney U test of Control with Visualization as grouping variable (H_{1d})

	Control
Mann-Whitney U	6876
Wilcoxon W	14136
Z	-0.618
Asymp. Sig. (2-tailed)	0.537

Table D.25: Mann Whitney U test of Challenge with Visualization as grouping variable (H_{1e})

	Challenge
Mann-Whitney U	6685.5
Wilcoxon W	13945.5
Z	-0.976
Asymp. Sig. (2-tailed)	0.329

Table D.26: Mann Whitney U test of Experience with Visualization as grouping variable (H_{1f})

	Experience
Mann-Whitney U	6602.5
Wilcoxon W	13862.5
Z	-1.138
Asymp. Sig. (2-tailed)	0.255

Table D.27: Mann Whitney U test of Active Learning with Natural Semantics as grouping variable (H_{1g})

	Active Learning
Mann-Whitney U	5946
Wilcoxon W	13206
Z	-2.412
Asymp. Sig. (2-tailed)	.016*

Table D.28: Mann Whitney U test of Interactive Learning with Natural Semantics as grouping variable (H_{1h})

	Interactive Learning
Mann-Whitney U	5961.5
Wilcoxon W	13221.5
Z	-2.368
Asymp. Sig. (2-tailed)	.018*

Table D.29: Mann Whitney U test of Authentic Problem with Natural Semantics as grouping variable (H_{1i})

	Authentic Problem
Mann-Whitney U	6364.5
Wilcoxon W	13624.5
Z	-1.596
Asymp. Sig. (2-tailed)	0.111

Table D.30: Mann Whitney U test of Control with Natural Semantics as grouping variable (H_{1j})

	Control
Mann-Whitney U	5988
Wilcoxon W	13248
Z	-2.311
Asymp. Sig. (2-tailed)	.021*

Table D.31: Mann Whitney U test of Challenge with Natural Semantics as grouping variable (H_{1k})

	Challenge
Mann-Whitney U	7066.5
Wilcoxon W	14326.5
Z	-0.253
Asymp. Sig. (2-tailed)	0.8

Table D.32: Mann Whitney U test of Experience with Natural Semantics as grouping variable (H_{1l})

	Experience
Mann-Whitney U	5632
Wilcoxon W	12892
Z	-2.988
Asymp. Sig. (2-tailed)	.003*

Table D.33: Mann Whitney U test of Active Learning with Interaction as grouping variable (H_{1m})

	Active Learning
Mann-Whitney U	6952
Wilcoxon W	14212
Z	-0.477
Asymp. Sig. (2-tailed)	0.633

Table D.34: Mann Whitney U test of Interactive Learning with Interaction as grouping variable (H_{1n})

	Interactive Learning
Mann-Whitney U	6504.5
Wilcoxon W	13764.5
Z	-1.33
Asymp. Sig. (2-tailed)	0.184

Table D.35: Mann Whitney U test of Authentic Problem with Interaction as grouping variable (H_{1o})

	Authentic Problem
Mann-Whitney U	6703.5
Wilcoxon W	13963.5
Z	-0.948
Asymp. Sig. (2-tailed)	0.343

Table D.36: Mann Whitney U test of Control with Interaction as grouping variable (H_{1p})

	Control
Mann-Whitney U	6488
Wilcoxon W	13748
Z	-1.357
Asymp. Sig. (2-tailed)	0.175

Table D.37: Mann Whitney U test of Challenge with Interaction as grouping variable (H_{1q})

	Challenge
Mann-Whitney U	6488.5
Wilcoxon W	13748.5
Z	-1.35
Asymp. Sig. (2-tailed)	0.177

Table D.38: Mann Whitney U test of Experience with Interaction as grouping variable (H_{1r})

	Experience
Mann-Whitney U	6910
Wilcoxon W	14170
Z	-0.553
Asymp. Sig. (2-tailed)	0.581

Table D.39: Mann Whitney U test of Active Learning with Immersion as grouping variable (H_{1m})

	Active Learning
Mann-Whitney U	6996
Wilcoxon W	14256
Z	-0.392
Asymp. Sig. (2-tailed)	0.695

Table D.40: Mann Whitney U test of Interactive Learning with Immersion as grouping variable (H_{1n})

	Interactive Learning
Mann-Whitney U	6997
Wilcoxon W	14257
Z	-0.388
Asymp. Sig. (2-tailed)	0.698

Table D.41: Mann Whitney U test of Authentic Problem with Immersion as grouping variable (H_{1o})

	Authentic Problem
Mann-Whitney U	7037
Wilcoxon W	14297
Z	-0.311
Asymp. Sig. (2-tailed)	0.756

Table D.42: Mann Whitney U test of Control with Immersion as grouping variable (H_{1p})

	Control
Mann-Whitney U	7015
Wilcoxon W	14275
Z	-0.353
Asymp. Sig. (2-tailed)	0.724

Table D.43: Mann Whitney U test of Challenge with Immersion as grouping variable (H_{1q})

	Challenge
Mann-Whitney U	7110
Wilcoxon W	14370
Z	-0.171
Asymp. Sig. (2-tailed)	0.864

Table D.44: Mann Whitney U test of Experience with Immersion as grouping variable (H_{1r})

	Experience
Mann-Whitney U	6956
Wilcoxon W	14216
Z	-0.465
Asymp. Sig. (2-tailed)	0.642

Vita

Hongbiao Yang received his B.S. degree in Logistics Engineering from Huazhong University of Science and Technology, China, in June 2009. After that, he worked as a game designer at G-bits Network Technology in Xiamen, Fujian province, China. Later on, he took the position as a software engineer in Huawei Technology in Shenzhen, Guangdong province, China. In 2012, he joined the Department of Industrial and Systems Engineering at the University of Tennessee, Knoxville. He also pursued a minor degree in Computer Science. His research interest focuses on applying Virtual Reality as a learning tool for hands-on skills training, and analyzing the corresponding learning effectiveness as well as providing guidance in Virtual Reality application design and development. He worked on various projects in the Natural Interaction Lab in the Department of Industrial and Systems Engineering.