12-10-2005

Reliability And Validity Of Virtual Build Methodology For Ergonomics Analyses

Tinghao Wu

RELIABILITY AND VALIDITY OF VIRTUAL BUILD

METHODOLOGY FOR ERGONOMICS ANALYSES

By

Tinghao Wu

A Thesis
Submitted to the Faculty of
Mississippi State University
in Partial Fulfillment of the Requirements
for the Degree of Master of Science
in Industrial Engineering
in the Department of Industrial Engineering

Mississippi State, Mississippi

December 2005

RELIABILITY AND VALIDITY OF VIRTUAL BUILD METHODOLOGY FOR ERGONOMICS ANALYSES

By

Tinghao Wu

Approved:

Vincent G. Duffy
Associate Professor of
Industrial Engineering
(Major Professor)

Mingzhou Jin
Assistant Professor of
Industrial Engineering
(Committee Member)

William N. Smyer
Associate Professor of
Industrial Engineering
(Committee Member)

Keith M. White
Adjunct Professor of
Industrial Engineering
(Committee Member)

Stanley F. Bullington
Professor,
Graduate Coordinator
Department of Industrial Engineering

Kirk H. Schulz
Dean of the College of Engineering

Name: Tinghao Wu

Date of Degree: December 9, 2005

Institution: Mississippi State University

Major Field: Industrial Engineering

Major Professor: Dr. Vincent G. Duffy

Title of Study: RELIABILITY AND VALIDITY OF VIRTUAL BUILD METHODOLOGY FOR ERGONOMICS ANALYSES

Pages in Study: 108

Candidate for Degree of Master of Science

This study was conducted to assess the validity and reliability of the Virtual Build methodology for ergonomics design and analysis. Thirty-six human subjects participated in this study and performed a set of six tasks. The tasks were performed twice in both real and virtual environment. The subject's motion in performing tasks was analyzed by ergonomics assessments by using Virtual Build methodology. Criteria-related validity was evaluated by comparing the Virtual Build ergonomic assessment results with manual calculation. Test-retest reliability was evaluated by correlating ergonomics assessment results between two trials.

The result shows that the Virtual Build methodology is reliable for ergonomic assessments. The Virtual Build with virtual environment has lower over-time reliability performance than the real environment. The t-test shows that the Virtual Build is valid for

1991 NIOSH lifting equation assessment when using real environment. Improvements need to be done to make Virtual Build valid when using virtual environment.

DEDICATION

I would like to dedicate this research to my parents, sister, niece, and beloved wife.

ACKNOWLEDGEMENTS

First, I would like to thank Dr. Vincent Duffy. He has always guided me in the right direction and opened his door to my questions. I would also like to thank my committee members, Dr. Mingzhou Jin, Dr. William Smyer and Dr. Keith White. Their comments have helped me throughout of all my work and added invaluable quality to my thesis.

I am grateful to my family members who have all helped get me to where I am today. Their understanding and encouragement supports me to proceed forward.

I would like to express special thanks to Amanda McAlpin, John McGinley, Renran Tian, Jihun Kang and Qiuli Yu, who contributed so much in this work. Thanks to all of the whole MSU IE and CAVS HFE members who added incredible experience to my life.

TABLE OF CONTENTS

CHAPTER Page

CHAPTER Page

LIST OF TABLES

TABLE Page

LIST OF FIGURES

FIGURE Page

CHAPTER I

INTRODUCTION

1.1 WMSD and Ergonomics

Ergonomics is the application of human physiology and psychology to the design of workplace (equipment and systems). One result of the absence of ergonomic consideration in the workplace can be the occurrence of Work-related Musculoskeletal Disorder (WMSD). WMSDs, such as low back pain, hand-arm vibration syndrome and carpal tunnel syndrome (CTS), account for a major component of the cost of work-related illness in the United States. Recent estimates of the costs associated with WMSD range from $13 billion to $54 billion annually (Panel on Musculoskeletal Disorders and the Workplace, 2001). The statistics from the annual survey of Occupational Injuries and Illness, conducted by the Bureau of Labor Statistics (1997), confirms that the problem is not only in health terms, but also in economic terms.

The National Institute for Occupational Safety and Health (NIOSH) recognizes the seriousness of the problem and collaborates with its partners from industry, academic lab and government interests, to group a team to evaluate and define the future research needs in the area of WMSD. This team developed a National Occupational Research Agenda (NORA) for WMSD in 2001(NIOSH, 2001). The NORA MSD agenda pointed out the most important research gaps in four primary topic areas, which include surveillance, etiology, intervention, and improving the research process. The highest priority for intervention research activity identified by the NORA team is to evaluate the effect of the following on development and prevention of WMSD:

1. Alternative (product and/or tool) design
2. Optimization of mechanical work demands (force, movement and posture) and temporal patterns of exposure
3. Emerging technologies

It is specified that researchers should investigate the work environment factors that affect posture, movement, force, exertion, and the interface between the worker and the equipment or the task. For the agenda of the improving the research process, researchers expressed frustration at the difficulties associated with gaining access to industrial sites to conduct research. An efficient approach to link the current research results to the actual workplace, instead of the “best case” scenario, will be a key progress.

1.2 Computer Aided Ergonomics

Ergonomics, the science to fitting the work environment to human worker, has received greater assists from computer-related technologies since the last two decades. Wells and Moore (1992) raised a framework for computer-assisted approaches to prevent WMSDs involving workplace design and modification which is shows in Figure 1. In this framework, the computer-aided technique address concerns like:

1. Can workers fit, reach and see?
2. Can workers avoid high external stresses?
3. Can workers avoid awkward postures?

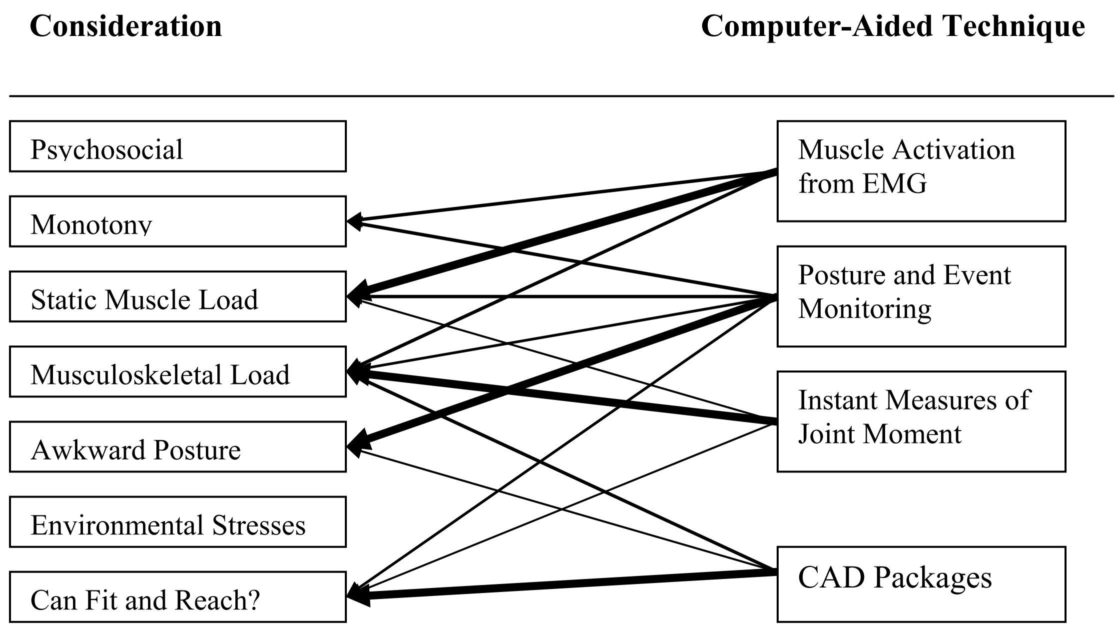

Figure 1 HIERARCHY OF ERGONOMICS REQUIREMENTS

This framework outlines the benefits of computer-aided techniques on the ergonomic research from different aspects, especially in the biomechanics-related area. It also points out that the ergonomics assessment of workspaces is obviously an important part of the design process, and ergonomic packages will rapidly be incorporated in CAD systems (Haslegrave et al., 1992).

1.3 Current Technologies

1.3.1 Digital Human Modeling (DHM)

The computer modeling of human was named “man-modeling” at its inception in the 1960s, but since the 1990s, it is now increasingly termed human modeling. The Center for Human Modeling and Simulation at University of Pennsylvania names it

"Virtual Human" and the Ergonomics Center at University of Michigan calls it "Digital Human Modeling" (DHM). In this study, DHM is used to represent the technology of using a computer to build a virtual representation of a real person to simulate human motion and exertions (Chaffin, 2001).

The DHM provides the ability to construct 2D or 3D human models from anthropometric data, which can be articulated between the body segments to simulate a wide variety of postures. These human models can then be used as substitutes for "the real human" in ergonomic evaluation of computer-based design for vehicle, work area, machine tool, assembly line, etc. (Badler, 1997). In conjunction with the CAD model of the product being designed, DHM enables conducting computer-based user trial to assess criteria such as fit, reach, vision, and the resulting constraints upon posture. Such predictions enable the ergonomist to be more proactive in the design process and to work closely with other design team members to achieve the ergonomic solutions to the design within the various financial, legal, engineering, and aesthetic constraints (Porter et al., 1999). The modeling has great benefits in testing alternative solutions to design problems, particularly where there are large differences in user anthropometrics or constraints on space, as is frequently the case in workplace design. Developing alternatives without the use of digital human modeling often requires expensive prototypes for iterative evaluation. Using DHM software, each new design can be simulated and analyzed on the computer without additional capital investment (Chaffin, 2002a; Rider et al., 2004).

Recent improvements in computation speed and control methods allow for the portrayal of 3D humans suitable for interactive and real-time applications. Over the last decade, several commercially supported human simulation programs have been licensed. Also, the Society of Automotive Engineers in North America and Europe has established a technical committee to define standards of DHM. The committee encourages the sharing of information and data to promote greater usage of DHM. Several automotive and aerospace companies are investing in further research and development for improving human models to allow their designers and engineers to perform ergonomics assessments of fit, clearance and sight line analyses prior to building prototype vehicles and workstation (Badler, 1997; Chaffin, 2003a). It is widely accepted that the DHM can assist us in designing better workplaces and products (Chaffin, 2001, 2002, 2003a; Porter et al., 1999; Gill et al., 1998).

Also, Chaffin (2002) pointed out that using human simulation within a digital mockup (DMU) would decrease the design time and enhance the number and quality of design options that could be rapidly evaluated by the design team. This is consistent with the concept of reducing the total design time and engineering costs by rapid prototype development and test, which will bring the economic benefits. (Badler, 1997; Chaffin, 2001, 2002; Gill et al., 1998). This view is represented in the Figure 2. With the reduction of total designs and engineering costs by using more computer-aided engineering (CAE) and DMU methods, we can achieve rapid prototype development and testing. And in some cases, human simulation is the only solution to verify that a design concept is acceptable to a prescribed population, since hardware prototypes, like the International Space Laboratory, are not available (Chaffin, 2001).

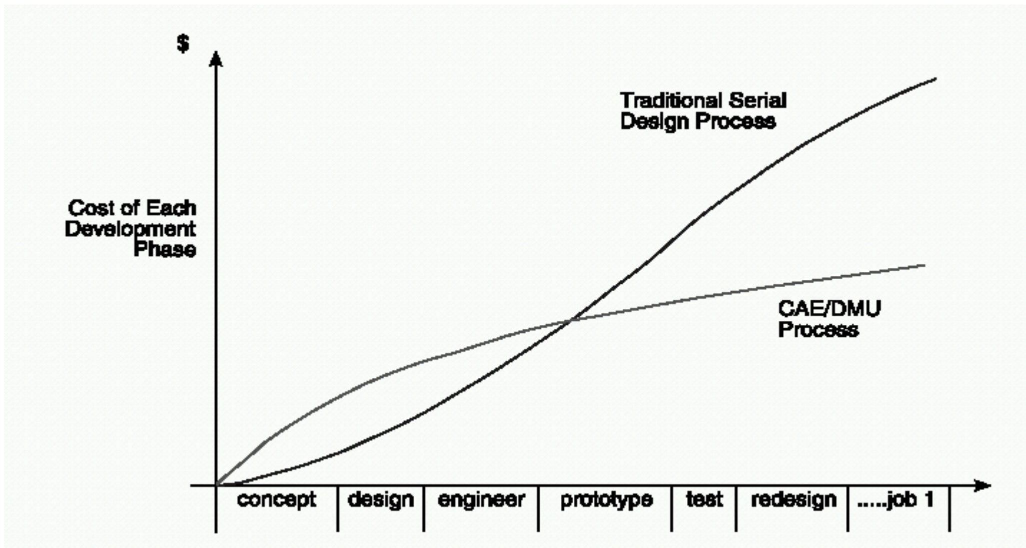

Figure 2 TYPICAL DEVELOPMENT PHASES AND COST PROFILES

1.3.2 Motion Capture (MOCAP)

The accurate and realistic representation of human behavior is one critical technical problem that should be solved before we can successfully implement the DHM into simulating human for ergonomics study (Rider et al., 2004), since workspace assessment covers more than the simple dimensions which are only related to static anthropometrics. The usefulness of modeling assessment would be determined by the capability of how far computer models are able to simulate human "behavior" (Haslegrave et al., 1992). Among the 7 case studies of Chaffin's book (2001), 5 of them pointed out that the "deriving postures or motion for dynamic analyses from motion capture files" is one of the major limitations of DHM technology. Chaffin (2002) also suggested that the biomechanical analyses of digital humans were dependent on accurate postures and movements, secondly only perhaps to the validity of the analysis tools used. Too often, the implementation of postures and motions is attempted through keyboard and mouse manipulation of the digital human model. Face validity of posture may be possible, but construct validity is difficult to establish (Chaffin, 2002, 2003a; Rider et al., 2004). Simulating human motion through keyboard and mouse manipulation is a tiresome and error-prone task. To overcome this problem and obtain accurate posture and movement, there are mainly two approaches. One approach is the inverse-kinematics model, which is built from the actual human motion data. The HUMOSIM lab in University of Michigan collected over 73,000 motions and functional regression analysis was used to predict the resulting movement based on the motion database (Faraway, 2000, 2001, 2003). Now, most DHM commercial software packages have implemented the inverse-kinematics model to make digital manikin behavior looks similar to actual human. Another approach is the motions capture technology. Motion Capture (MOCAP) is an attractive method for creating the movement for computer simulation of human action because it can provide realistic motion, which contains the nuance and specific details of particular performers (Gleicher and Ferrier, 2002).

There are three main kinds of motion capture systems based on the mechanism of tracking targets. They are optical-based MOCAP, magnetic-based MOCAP and mechanical-based MOCAP. Each kind of motion capture system has its own advantage and disadvantages. Generally, the optical-based MOCAP system has the highest capture rate and also the highest price. The mechanical-based MOCAP system has the biggest capture volume and less environment limitations. Both the optical-based and magnetic-based MOCAP systems have more requirements on the environment than the mechanical-based MOCAP system does (Delaney, 1998).

It is recognized that the accuracy of the motion capture system is affected by the following factors:

1. Marker movement
2. Sensor noise
3. Restriction on environment
4. Frame rate

Further, the calibration of the motion capture system has a significant effect on the overall performance. A good calibration is the basis for all motion capture work.

1.3.3 Virtual Environment (VE)

Virtual Environment (VE) is refereed to the 3D data set describing an environment based on real-world or abstract objects and data (Stanney, 2002). As the new generation concept for the Human Machine Interface (HMI), VE has been largely used to create interactive virtual world. With current development of computer technology, VE has been widely used for the design and evaluation of future products and processes. VE can provide accurate and realistic representation of the real workspace. Wilson et al. (1996) listed a number of areas where companies could benefit from VE. These areas include: job training, work aids, visualization and communication aid, testing human-machine interfaces, and a safe alternative to reality. Simulating objects or environments allows testing at early stages of development, thus reducing the guesswork and ensuring a quality product. More and more attention in ergonomic fields has been given to VE (Cerney et al., 2003, 2002; Davies, 1997).

Besides those academic researches, there is a marked rise in recent years in the number of commercial human mannequin packages, which have capabilities to build the VE environment and place the digital human model inside the VE for different purpose studies. Some of these packages, such as RAMSIS, UGS JACK, and SAFEWORKS, are designed for human factors and ergonomics assessment. These new VE packages endow the virtual human model with anthropometric and biomechanical data, not to mention physiological libraries encompassing energy expenditure, psychomotor parameters and so on. The use of ergonomic knowledge to bring a new methodological credibility to many engineering projects based on VE is also growing steadily (Stone, 2002).

There are two main kinds of VE. The usual definition of VE involves full immersion. That is, the user wears head-mounted stereo displays to provide full visual immersion and special gloves that allow six-degree-of-freedom input for directly manipulating the environment. We call it "Immersive VE". The parallel to "Immersive VE" is "Non-Immersive VE", which exposes the virtual world to human by means of conventional graphics workstations using a monitor, a keypad and a mouse. In Non-Immersive VE, the scene is displayed with the same 3D depth curs used in Immersive VE: perspective view, hidden-surface elimination, colors, texture, lighting, shading and shadows (Roberston, 1993). Full immersion is often seen as a major advantage. But the previous studies suggest that, for many applications, the same effect is possible with proper 3D cues and interactive animation (Robertson et al., 1993). Most advantages of

Non-Immersive VE, and also disadvantages of Immersive VE, are technical-related (Roberston, 1993). Some of them, such as display jitter, time lag in six-degree-of-freedom input devices, and display resolution, have been improved with the rapid progress of computer techniques.

1.4 Solution

1.4.1 Virtual Build

A structure called "Virtual Build" was proposed by Ford Auto Company in 2003. The Virtual Build integrates the DHM, MOCAP and VE for ergonomic research (Brazier et al., 2003). Virtual Build is a systematic methodology for future proactive engineering or concurrent engineering concept.

If we take Virtual Build as a black box system, the inputs are:

1. Environment (real or virtual)
2. Population information
3. Descriptive parameters of the task that is going to be analyzed.
4. Human motion of interaction with the environment

The systematic structure of Virtual Build is showed in Figure 3, which was described by Chaffin (2005).

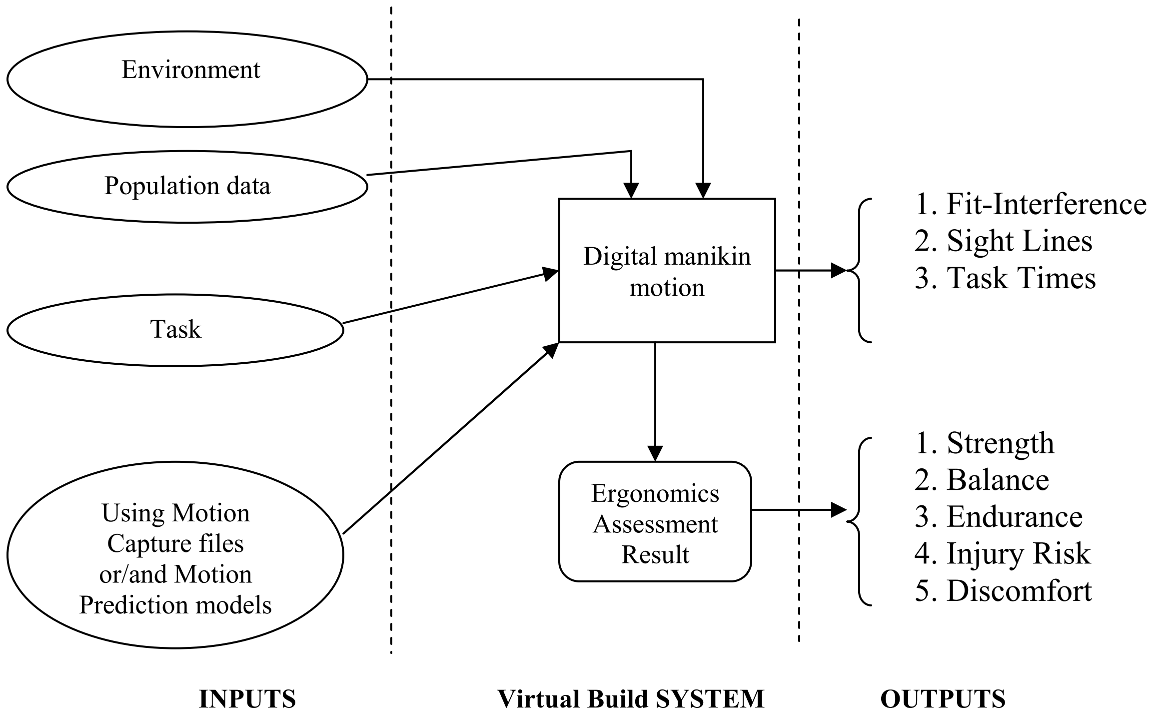

Figure 3 SYSTEMATIC VIEW OF VIRTUAL BUILD

In this system, the environment may be a physical mockup or virtual environment, with which that human can interact. The mockup or virtual environment should actually represent the real workspace or workstation. Human motion can either come from the motion capture system, from the motion prediction model or even from manual setup. Virtual Build also takes the anthropometric information as one input to set up the digital human modeling. In order to perform the ergonomic assessment, corresponding task-related descriptive parameters need to be input into the system. These descriptive parameters include information like external loading, work frequency, etc.

The following figure shows the system components and the general integration structure of “Virtual Build” methodology with virtual environment setup. In Figure 4, the dark solid line represents physical connection and dash line represents system internal features. The arrow shows the direction of information flow. The motion capture system tracks the human subject's activity, and creates MOCAP marker model based on the human real motion. The motion capture system is connected with DHM & ergonomics analysis system. The MOCAP marker model is transferred to DHM & ergonomics analysis system to animate the digital human model, so that the digital human model can simulate the actual human subject. Then the DHM & ergonomics analysis system can conduct the ergonomic assessments based on the digital human model. The VE system provides the virtual view with which the human subject interacts.

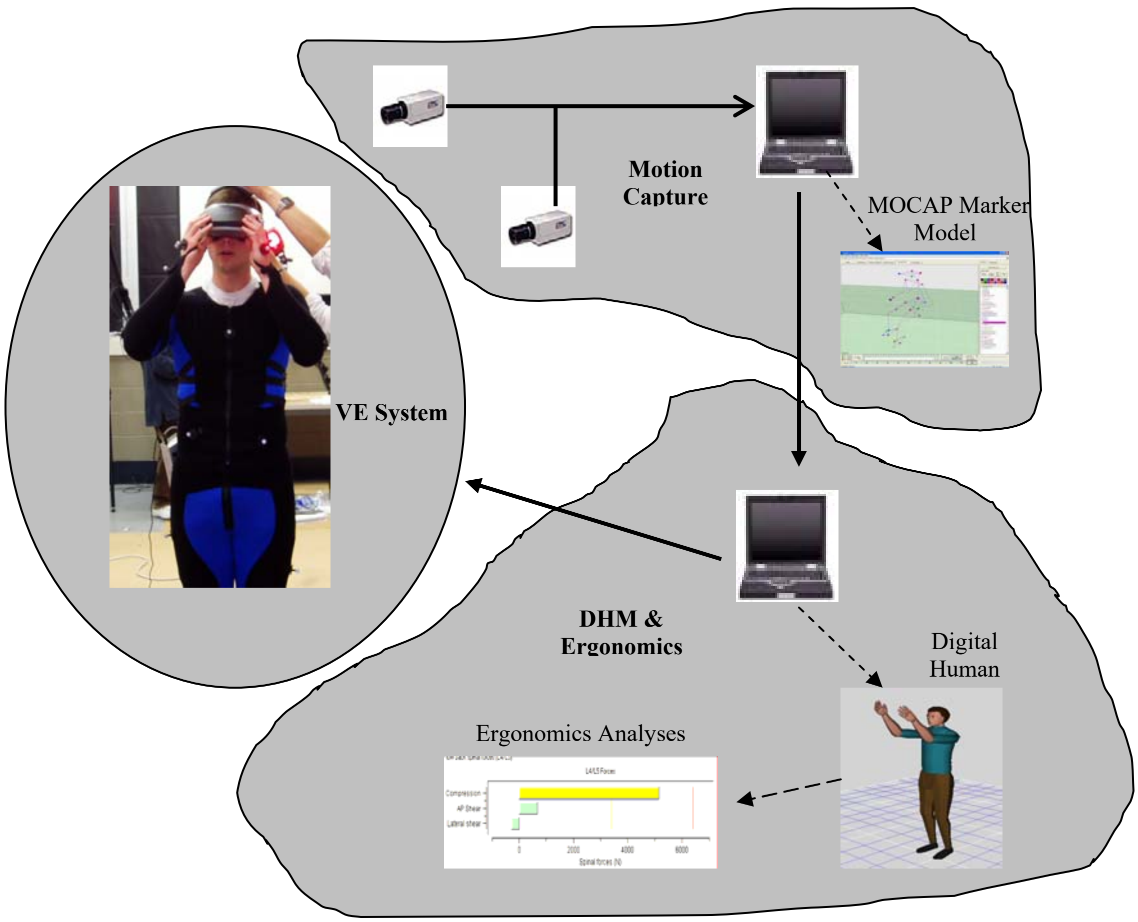

Figure 4 INTEGRATION STRUCTURE DIAGRAM OF VIRTUAL BUILD

At certain situations, physical mockup is used instead of virtual environment. At this case, there is no VE system. Human subject interacts with a physical mockup, which represents the workstation. MOCAP system will track human subject's motion, and it then streams motion data into DHM & ergonomics analysis system for ergonomic assessments.

Based on the different integration setup, most previous Virtual Build-based ergonomic studies can be categorized into following three types.

1. DHM Simulation
2. DHM + MOCAP + Mockup
3. DHM + MOCAP + VE

In DHM Simulation, digital manikin's posture and motion are implemented through either manual setup or inverse-kinematics. By observing the actual operator's activity and then simulating by keypad input or mouse drag, the manual setup has little construct validity and is difficult to control. Meanwhile, this task is tiresome. Inverse-Kinematics implements the motion to digital manikin by digitizing the actual activity into data and then using statistics to build mathematic model to predict other motions. Inverse-Kinematics has higher construct validity than manual setup (Faraway, 2001, 2000, 2003). However, the Inverse-Kinematics model has several problems:

1. The reliability of the model
2. The prediction error of the model
3. It asks for motion capture first, to collect the baseline data. Without motion data, through motion capture or not, no inverse-kinematics model can be built.

Because MOCAP technology provides the accurate motion data, which is recorded from actual human movement, it has higher facial validity than manual operation in simulating human motion for human modeling purpose. While the integration of the DHM with MOCAP demonstrates many advantages, it is also restricted by the limitation of motion capture systems. Most motion capture systems have certain requirements for the space volume. For example, the optical-based motion capture system requests a space without obstacles and reflective objects, and the magnetic-based motion capture system requests a space free of any metal objects, which may affect the magnetic field. Therefore, most previous studies were conducted in well-controlled lab environment. To facilitate the motion capture system and also save costs, physical mockups are used to represent the work environment. There is a concern of the validity of the mockup environments. How well can it represent or simulate the real workspace? Will the human behavior be the same in the mockup environment as in the real workspace? In some situations, we cannot even build a mockup, so the VE is introduced into the integration to facilitate testing when a full physical mockup is not available or possible.

The integration of DHM, MOCAP and VE provides a theoretical sound solution for the ergonomics study in designing a future factory or redesigning an existing workspace. With the CAD data, one can build the virtual environment for the work station that needs to be studied, and through exposing human into this virtual environment, motion capture can record the details of working. Then motion data can be imported into digital human modeling systems to conduct the ergonomics assessments. We can then derive the efficient assessment result.

Table 1 lists some studies that have been done using the Virtual Build methodology.

Table 1. LITERATURE OF USING VIRTUAL BUILD FOR ERGONOMICS

Author	Methodology	Purpose
Ford Company (Brazier, 2003)	DHM + Magnetic MOCAP	Vehicle Design
Ford Company (Brazier, 2004)	DHM + Magnetic MOCAP + VE	Auto assembly line Design
Univ. of Michigan (Kevin et al., 2004)	DHM + Inverse Kinematics	Maintenance work design + vehicle Design
Miss. State Univ. (Li et al., 2004)	DHM + Optical MOCAP + VE	Design justification

1.4.2 Concerns

While being increasingly used, the Virtual Build methodology has not been studied regarding its validity and reliability for ergonomic research. The validity and reliability of Virtual Build are key functions that must be addressed before it can be fully implemented. Validation is an essential part of the development process. Without it there is no way to ensure that the method actually captures representative activity of the operational environment (Stanton et al., 1997). In a review of ergonomic methods, Stanton and Young (1995) identified over 60 methods available to ergonomist. However, despite the proliferation of methods, few attempts have been exerted to validate them. In a critical paper, Kanis (2000) identified the uses and abuses of validation in ergonomics research. He argued that either validation studies were simply not undertaken, or they are undertaken with inappropriate methods.

The Virtual Build methodology must be tested for its validity and reliability before it can be widely accepted in ergonomic analyses. The validity must to be proven before we accept DHM as a useful approach (Badler, 1997; Peters et al., 2002). Chaffin (2003a) proposed that the topic of model validation methods was a challenge faced by the DHM technology. Badler (1997) pointed out that the fidelity to human size, capabilities, and joint and strength limits were essential to applications such as ergonomics or design evaluation. One of the most pertinent questions asked about DHM concerns the accuracy and dependability of the results from an analysis, compared to conclusions from a manual assessment. The question of accuracy and dependability becomes important when the technology is used to generate specific value for the analysis of design. Mital, et al. (1996) investigated a manual lifting task and made a comparison between the manual and software result for the problem. He found that there is a significant difference between the NIOSH lifting equation result from the manual and software result. Haslegrave, et al. (1992) pointed out the validity and usefulness of modeling assessment will only be as good as the human models contained in the computer-aided package. The ease of using DHM and accuracy of the resulting simulation determine the acceptance of this technology. For validity reasons, Porter (1999) suggested that the DHM should not replace user trials with full-size mock-ups.

Validity of the motion capture system is a topic that arouses the attention of both the vendor companies and the users. Every company claims its motion capture system has a higher accuracy, and every user wants to find out the most accurate system within the budget. For the motion capture system, the noise and deviation represent the random and non-random errors of the result from the reality and can be used as measures for the accuracy for the system. Ehara, et al. (1995, 1997) tested 11 commercially available 3D optical-based motion capture systems to evaluate performance in a clinical gait study. The accuracy and the noise of all those 3D motion capture systems were evaluated using a well-designed experiment. The result shows that optical-based motion capture systems can control the error around several millimeters with small amount of noise (Standard Deviation). The result shows that the 3D optical-based motion capture system can accurately track the reflective markers motion in the capture volume.

VE technology allows the computer users to enter the computer-generated virtual world and interact with graphical objects and virtual agents with the sense of reality. The validity and fidelity of a Virtual Environment involve the degree to which the VE duplicates the appearance. The fidelity is a concept used to measure human perceptual response to the environment. As humans lose some perceptions in the VE, compared to real environment, some assistant feedback will be helpful to improving the human's performance in the virtual environment. The VE with feedback, like the collision detection and hybrid Immersive-VR (Lok et al., 2003), can provide additional sense of reality to human, which leads to higher task performance. In order to increase the sense of reality for a VR system, it is necessary to detect collision between the graphic objects in real time (Youn et al., 1993).

CHAPTER II

VALIDITY AND RELIABILITY

2.1 Random and Nonrandom Error

There are two basic kinds of errors that affect measurements: random and nonrandom error. Random error is the term used to designate all of those chance factors that confound the measurement. The amount of random error is inversely related to the degree of reliability of the measurement. Thus, a highly reliable measure is one that leads to consistent results on repeated measurements because it does not fluctuate greatly due to random error. The nonrandom error has a systematic biasing effect on measuring instruments. Nonrandom error lies at the very heart of validity. The invalidity arises because of the presence of nonrandom error. The validity depends on the extent of nonrandom error present in the measurement process (Carmines, 1979). Carmine's definition of validity and reliability is represented in the following figure.

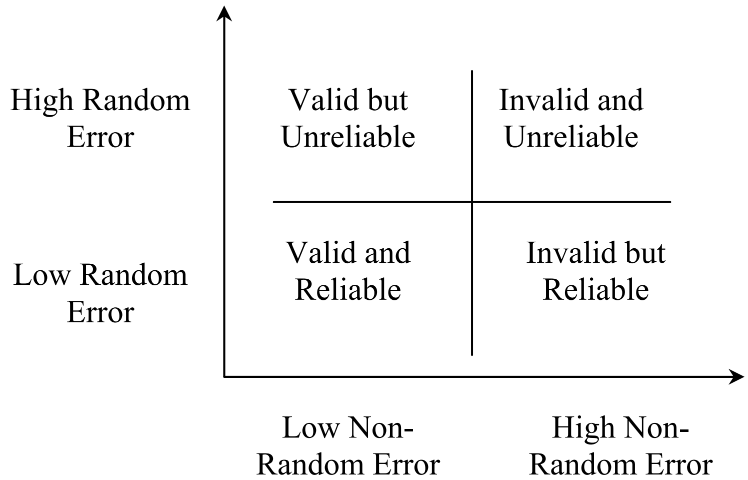

Figure 5 VALIDITY AND RELIABILITY

Right now, there is a trend to take the reliability an imperative requirement for having validity. A method can be reliable, but not valid. But in order for a method to be valid, it must be reliable (Fagarasanu, et. al., 2002; Hager, 2003). A more strict definition of validity is represented in Figure 6. The random and nonrandom error figure (Figure 5) can be used to help us understand the causes of a method's characterization as invalid or unreliable. It can also help us to identify the possible methods of improving invalidity and reliability.

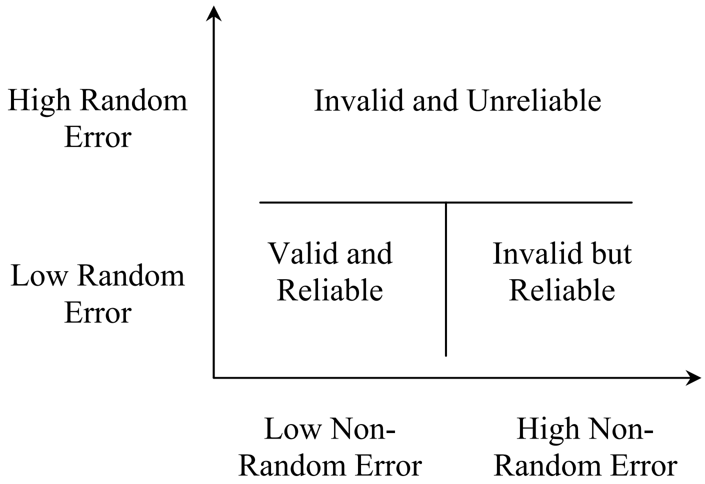

Figure 6 VALIDITY AND RELIABILITY

2.2 Reliability

For an experiment or measure, good reliability implies that it has repeatable results. The Merriam-Webster Dictionary defines reliability as “the extent to which an experiment, test, or measuring procedure yields the same results on repeated trials”. Reliability involves the precision of methods and the level of credibility placed on results (Aarass et al., 1996). Reliability is important, because issues such as measurement error and subject variability can have a negative impact on statistical results and interpretation of these results. The concept of reliability is different from validity. A method is valid if it measures what it is designed to measure. A method is reliable if it can get same result in repetitive tests. A method can be reliable, but not valid. But in order for a method to be valid, it must be reliable (Fagarasanu et al., 2002). The reliability will affect level of validity. To assess reliability, there are several methods, which include split-halves method, alternate-form method, internal consistency, inter-rater method, and test-retest method.

1. Test – retest: Administer instrument, analyze responses, administer instrument again under exact conditions, and correlate responses.
2. Alternative forms: Administer instrument, analyze responses, administer different form of same survey, and correlate response.
3. Split halves: Administer instrument, separate responses by odd and even items, and correlate two halves.
4. Internal Consistency: Administer instrument, computer inter-item correlation matrix, computer mean inter-item correlation.

Freivalds gave several reliability index which would be suitable for categorized data, such as survey questionnaire. Freivalds also pointed out that for continuous data measurement, reliability index based on statistical approaches is more proper (Freivalds, 2004). This research focuses on the test-retest reliability, which measures consistency over time. An example of test-retest reliability involves the subject taking the same test during two different times. It is generally thought that test-retest is more costly than the others, but it is a simple and clear reliability method (Hager, 2003). For our purpose to test the over-time reliability and identify the possible factors to affect the reliability of the integration of DHM, MOCAP and VE, test-retest reliability fits the requirement very well.

2.2.1 Reliability Indexes

A number of indexes for reliability are available. The literature provides much conflict on which index is most appropriate for use. For continuous data, consensus is measured as the Pearsonian correlation between the ratings for pairs of raters. A literature review of recent research showed that even though Pearson's correlation coefficient (*r)* was commonly used in the past studies, it is not an acceptable form of measuring test-retest reliability (Denegar et al., 1993; Keating, 1998). It is reported, for small samples (<15), Pearson's r overestimates test-retest correlation (Bland et al., 1986; Larssona et al., 2003). Intra-class Correlation (*ICC*) is preferred when sample size is small (< 15) or when there are more than two tests (one test, one retest) to be correlated.

ICC is the ratio of the between-subjects variance divided by the total variance (Denegar et al., 1993). Following Ebel (1951): Let A be the true variance in subjects' rating due to the normal expectation that different subjects will have true difference scores on the rating variable. Let B be the error variance in subjects' rating attributable to inter-rater unreliability. The intent of *ICC* is to form the ratio:

$$ICC = \frac{A}{B}$$

That is, Intra-Class Correlation (*ICC*) is to be true inter-subject variance as a percent of total variance, where total variance is true variance plus variance attributable to inter-rater error in classification. B is simply the mean-square estimate of within-subjects variance computed in *ANOVA. ICC* will approach 1.0 when mean-square estimate of between-subjects variance close to 0, that is, when there is no variance within targets, indicating total variation in measuring on the scales is due solely to the target variable. For instance, one may find all raters rate an item the same way for a given target, indicating total variation in the measure of a variable depends solely on the values of the variable being measured, that is, there is a perfect inter-rater reliability. *ICC* is 0 when within-group variance equals between-group variance, indicative of the grouping variable having no effect. Though less common, the *ICC* can become negative when the within-group variance exceeds the between-group variance. In this situation, it takes as no reliability as *ICC* equals to 0.

There are six forms of *ICC*, which are labeled *ICC (1, 1), ICC (2, 1), ICC (3, 1), ICC (1, K), ICC (2, K), ICC (3, K)* (Shrout et al., 1979; Shrout 1999). Test-retest reliability with a single rater and a 2-way ANOVA is estimated in following equation.

$$ICC(2,1) = \frac{BMS - EMS}{BMS + (k-1)EMS + k[(TMS - EMNS)/n]}$$

* BMS = Between-subject mean squre
EMS = Error mean square
TMS = Trial mean square
K = number of trials or evaluators
N = number of subjects.

ICC can range from 0 to 1; where 0 indicate no reliability and 1 indicate perfect reliability. A negative *ICC* indicates that the within-subject variance exceeded the between-subjects variance and is equivalent to an *ICC* of 0, or no reliability.

2.2.2 Classification of *ICC*

With the availability of reliability scale, the *ICC* index, how to classify the level of reliability is often a disagreement among different studies. Researchers tend to use descriptions, such as perfect, good, poor and etc, to associate with range of reliability. But there are disagreement about the range and association.

In some studies, *ICC* ranging from 0.7 to 1 suggests that there is a good/high/ excellent correlation between classes (Wertheim et al., 2004; Dankaerts et al., 2004; Henriksen et al., 2004; Osman et al., 2004). While some other studies (Flisher et al., 2004; Parkinson et al., 2004; Tzannes et al., 2004) define the *ICC* range as perfect (0.8 to 1), substantial (0.6 to 0.8), moderate (0.4 to 0.6) and poor (0 to 0.4). Following table summarizes the commonly representation of *ICC* in several ergonomics, psychology and medical studies.

Table 2. LITERATURE OF CLASSIFICATION OF *ICC*

Literature	ICC Range and description
Flisher et al., 2004	0 – 0.4 poor , 0.4 – 0.6 moderate, 0.6 – 0.8 substantial, 0.8 – 1 perfect,
Henriksen et al., 2004	0 – 0.4 poor , 0.4 – 0.7 fair, 0.7 – 1 good,
Koumantakis et al., 2002	0 – 0.69 poor, 0.7 – 0.79 fair, 0.8 – 0.89 good, 0.9 – 1 high
Stokdijk et al., 2000	0 – 0.39 poor, 0.4 – 0.59 fair, 0.6 – 0.74 good, 0.75 – 1 excellent
Shrout 1998	0 – 0.1 none, 0.11 – 0.4 slight, 0.41 – 0.6 fair, 0.61 – 0.8 moderate, 0.81 – 1 substantial
Bartko et al., 1996	0 – 0.6 poor, 0.6 – 0.8 good, 0.8 – 1excellent
Fleiss 1986	0 – 0.4 poor, 0.4 – 0.75 fair to good, 0.75 – 1excellent
Landis et al., 1977	0-0.2 slight, 0.2 – 0.4 fair, 0.4 – 0.6 moderate, 0.6 – 0.8 substantial, 0.8 – 1 perfect

In this study, we define the *ICC* range as following table:

Table 3. CLASSIFICATION OF *ICC*

ICC Range	Meaning	Notes
(0.80 , 1.0]	Excellent	Perfect match
(0.60 , 0.80]	Good	Relative high agreement
(0.40 , 0.60]	Moderate	Though reliability not high, but possible being improved
[0.00 , 0.40]	Poor	No or few correlation

For "Excellent" and "Good" reliability, the two tests correlate with each other very well and they are expected for the ideal reliability test. For the Moderate reliability, there are possible spaces to improve the reliability through methods, such as increasing sample size; even through the reliability is not high. Poor reliability indicates that reliability is quite low that it would not be very useful or applicable.

In this study, we follow the methodology of Yeung (2002) to test the validity and reliability. Validity is evaluated through comparing the experiment result with the criteria, the manual calculated ergonomic assessments. And test-retest reliability is evaluated through correlating the experiment result in two different times. Also, with the comparison between different integration degrees, we may identify the possible factors to affect the validity of Virtual Build methodology.

2.3 Validity

In general sense, a measure is valid if it does what it is intended to do. The indicator of measure's validity is the extent that it measures what it purports to measure (Carmines, 1979). And validity is a matter of degree, not an all-or-none property.

To measure validity, there are four methods, which are construct validity, content validity, and criterion-related validity and face validity.

1. Construct validity: is the underlying construct or theoretical foundation of the method consistent with research and information on this topic.
2. Content Validity: Does the content of the item in the instrument accurately reflect the underlying construct?
3. Criterion-related validity: Does the method contain the proper criteria for measuring the traits or constructs of interest?
4. Face validity: Does the method look like it will measure what it is supposed to measure?

Validity of one measurement refers to the accuracy of a measure. A valid measurement should be close to what it intend to measure within an acceptable error limits. The validity of the measure is usually estimated by the size of their correlation (Carmines, 1979). In practice, for some well-defined group of subjects, one correlates performance on the test with performance on the criterion variable (Yeung et al., 2002). This correlation, for obvious reasons, is sometimes referred to as a validity coefficient. The test will not be useful unless it correlates significantly with the criterion. The higher the correlation, the more valid is the measures for this particular criterion.

Nunnally(1978) argued that even modest correlations (e.g., a correlation of .30) between test and criterion can prove quite useful for selection purpose.

2.4 Discussion

In his book, Freivalds (2004) pointed out that a valid ergonomic assessment tool will allow the ergonomist to make useful inferences about an individual working on a given job. Freivalds (2004) also pointed out that a valid ergonomic assessment tool needs to be reliable.

Virtual Build is not an ergonomic assessment tool. It is not Virtual Build's purpose to identify the ergonomic risk factors. Virtual Build provides the channel to use different ergonomic assessment tools to identify the ergonomic risks. To evaluate the validity and reliability of Virtual Build, we use some well-established and accepted ergonomic assessments.

CHAPTER III

RESEARCH OBJECTIVE

3.1 Objective

The purpose of this study is to test the validity and reliability of the Virtual Build methodology for ergonomic assessments, and also, identify the possible factors that may affect the validity and reliability.

The Virtual Build methodology shows promising for ergonomic study, especially in the proactive design process. Ergonomic studies have been done using this methodology, fully or partially, without validating this methodology. There are some studies about the accuracy of DHM, MOCAP or VE, but there is no research has been done to test the overall validity of Virtual Build methodology for ergonomic assessments. Unlike the movie animation, the ergonomic study asks for an accurate description of the human interaction with the environment. The validity is very important for ergonomic researches. The validity decides the acceptance of this methodology in ergonomics field. This study is conducted to test the criteria-related validity of the Virtual Build methodology. The manual measured and calculated ergonomic assessment is treated as the error-free criteria. It is expected that the Virtual Build ergonomic assessment is not significantly different from the criteria, and then prove the validity of the new methodology.

Few studies have been done to test the reliability of the DHM, MOCAP, and VE. Reliability is important, because if reliability is not account for, statistical results can be misinterpreted. Lower reliability of measures may negatively affect the validity of measures. This study focuses on the over-time reliability of the Virtual Build methodology for ergonomic analyses. A test-retest experiment study is designed to evaluate the over-time reliability.

The result of this study can be used to determine if the Virtual Build methodology is acceptable for ergonomic research. High validity and excellent reliability are expected for introducing the methodology to ergonomics field. Poor or fair reliability implies that there are some random factors with significant effect on the result. Standardizing those factors can reduce the random error, and may improve the reliability performance. Poor or fair validity implies that there are some systemic factors with significant effects. Those systemic factors need to be removed before the methodology can be fully accepted.

3.2 Questions

This study does not cover random factors, like motion capture system setup/calibration and reflective markers placement. These factors are left for future study.

Following questions needs to be answered by this study.

1. Can the Virtual Build methodology provide high criteria-related validity?
2. Can the Virtual Build methodology provide excellent test-retest reliability?
3. Is there any significant difference in performance between different Integration Levels of Virtual Build methodology?

CHAPTER IV

METHODOLOGY

4.1 Subject

A total of 36 human subjects were invited to participate in this study. All subjects were recruited form Mississippi State University campus after the screen of musculoskeletal disorder and simulator sickness.

4.2 Instruments

An 8-camera optical motion capture system, manufactured by Motion Analysis Company, was setup with 60 frames per second to track and record the motion. The motion capture system was calibrated on the volume with length of 3m, width of 2m and height of 2.5 m. The captured motion data was saved and also streamed into UGS JACK system, which was used to create the digital manikin model, to perform the ergonomic assessments, and also to generate the virtual environment. A 5DT Head-Mounted Display 800 with resolution of 800X600 was connected with UGS JACK system to expose the virtual environment to subjects. Following picture describes the instrument setup.

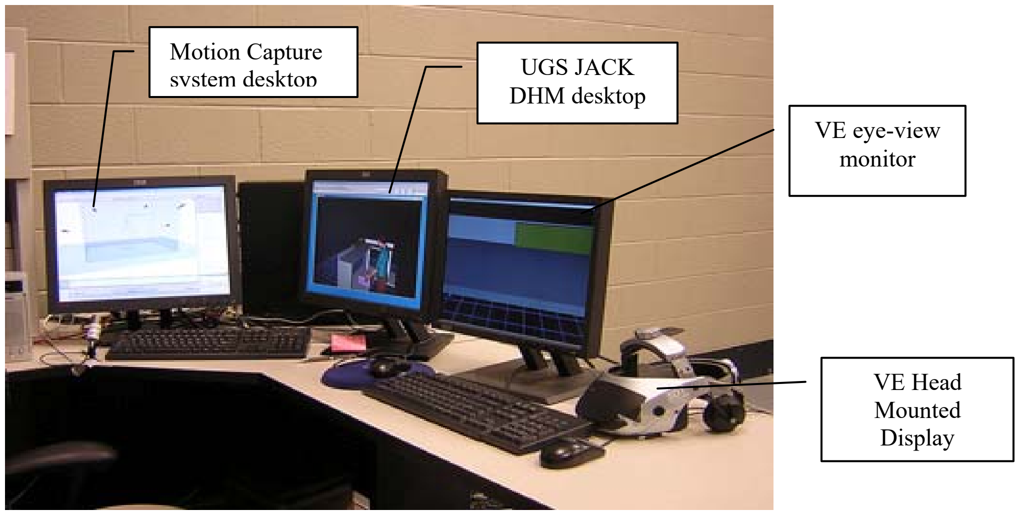

Figure 7 EXPERIMENT INSTRUMENT SETUP

4.3 Terms

The following terms are frequently used in this paper: Integration Level (IL), Task (TA), Trial (TR), Anthropometric Inputs (AI), and External Loading (EL)

4.3.1 Integration Level (IL)

Integration Level (IL) is the different degree of integrating DHM, MOCAP and VE. There are 3 levels of integration:

1. I: DHM Simulation
2. II: DHM + MOCAP + Physical Mockup
3. III: DHM + MOCAP + VE

No human subject is involved in the Integration Level I experiment. Integration Level I does not include MOCAP and VE. In Integration Level I, digital manikin is manipulated by a researcher through using Inverse-Kinematics models. In level II and III, human subject interacts with physical mockup and virtual environment respectively, the actual motion data drives the digital manikin in DHM system.

4.3.2 Task (TA)

Task (TA) is the different job that human subjects/digital manikin perform in this study. Following list is the tasks in this study.

1. A Front Lifting (FL): with External Loading : 0 lb , 1 lb, 20 lbs
2. An Side Lifting (SL) : with External Loading : 1 lb
3. A Forward Reaching (FR)
4. A Horizontal Pushing (HP)

Following pictures shows details of each task. All lifting tasks ask the human subject/ digital manikin to pick a box up from the table surface in front of him/her.

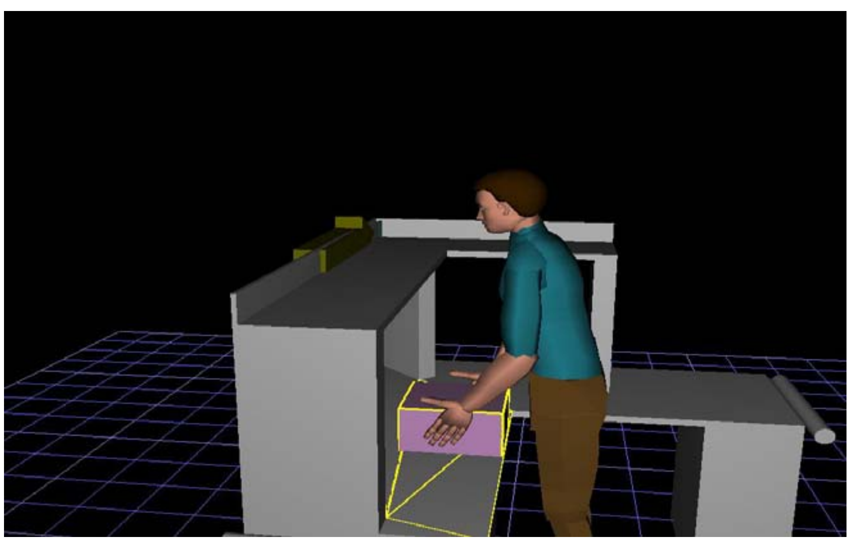

Figure 8 ORIGINAL POSITION OF LIFTING

The front lifting task asks human subject/ digital manikin place the box on the shelf platform in front of him/her. Following pictures shows the human at the destination position of front lifting task.

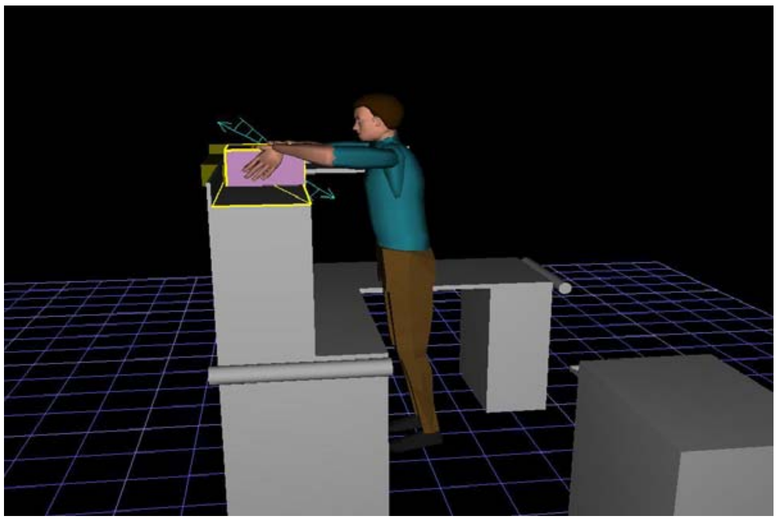

Figure 9 DESTINATION POSITION OF FRONT LIFTING

The side lifting task starts from the same original position as the front lifting task, but its destination is at the 90 degree right from that of front lifting task. Following picture is the human subject at the destination position of side lifting task. Subject twists his/her body to finish the side lifting task. The feet do not move during the side lifting.

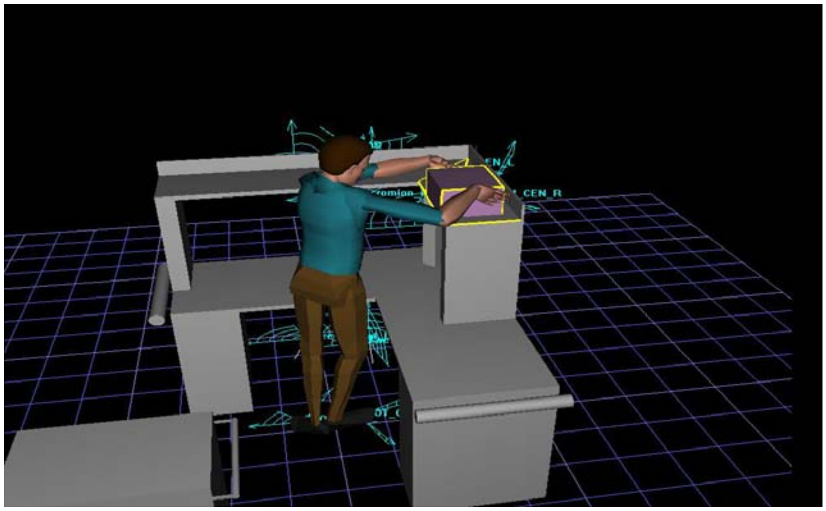

Figure 10. DESTINATION POSITION OF SIDE LIFTING

Front reaching task requires human subject/digital manikin to extend his/her right upper extreme to touch the corner of upper shelf.

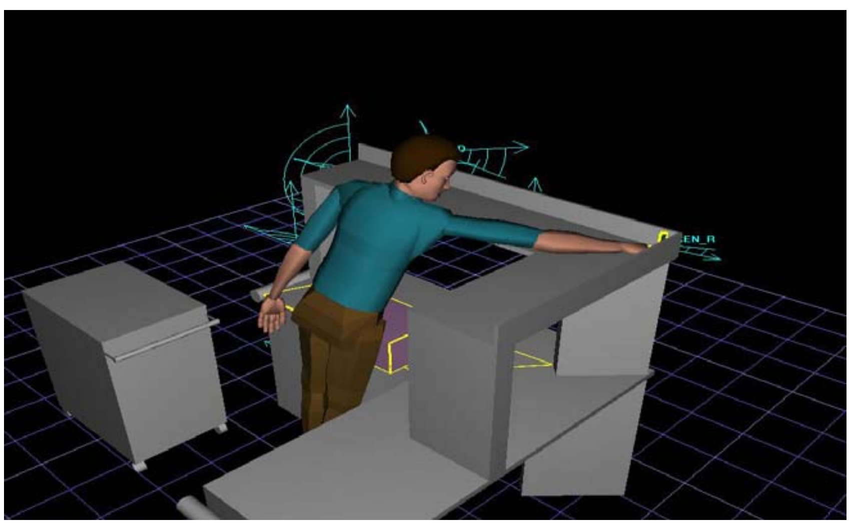

Figure 11 REACHING

The pushing task requires subject / digital manikin to place his/her hands on the handle of a tool cart and push the cart move forward.

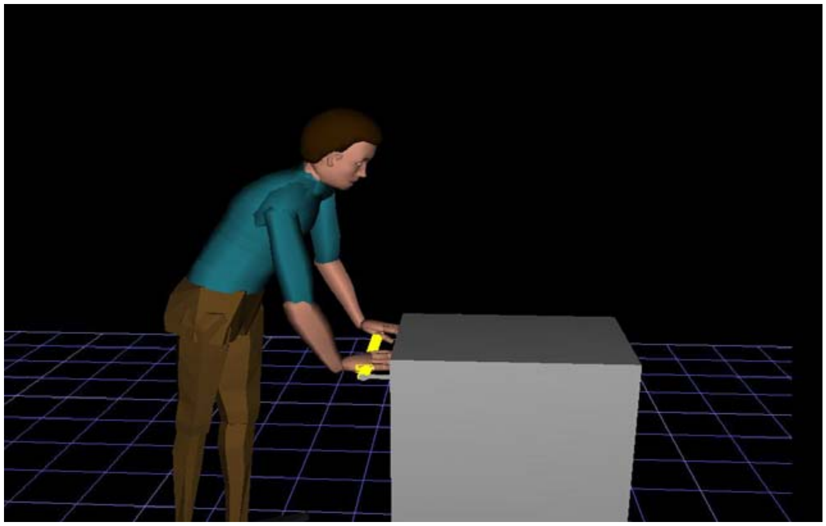

Figure 12 PUSHING

4.3.3 External Loading (EL)

For front lifting task, there are three different loading weights, 0 lb, 1 lb and 20 lbs. Subjects perform all three loadings lifting. The 0 lb lifting is to simulate the lifting with hand emptied. The 1 lb lifting is to lifting an empty box weighted 1 lb, and the 20 lbs lifting is to lifting the box with extra loading insides and weighted 20 lbs in total. So The External Loading (EL) has three levels.

1. 0 lb: empty-hand lifting
2. 1 lb: empty-box lifting
3. 20 lbs: loaded-box lifting

4.3.4 Trial (TR)

Trial (TR) is the number of times that the human subject/digital manikin performs each task. For Integration Level I, the TR is the number of times that the digital manikin is manually manipulated for each task. For Integration Level II and III, the TR is the number of times that each subject performs each task in each Integration Level. There are two trials of each task in each Integration Level. TR is 2 for all tasks.

4.4 Independent Variables

4.4.1 Front Lifting Task

For the front lifting task, independent variables include Integration Level (3 levels), External Loading (3 levels). All human subjects perform 2 trials of front lifting task with all three External Loadings in both real and virtual environment (Integration Level II and III), totally 12 front lifting trials. In Integration Level I, researchers create the UGS JACK digital manikin with the actual anthropometric size information, and then setup the External Loading parameters (3 levels) in UGS JACK and manipulate the human manikin model in UGS JACK to finish 2 trials of the front lifting task for Integration Level I. There are 6 manipulation of the digital manikin with the actual anthropometric size information.

Following table shows the design of experiment for front lifting task.

Table 4. INDEPENDENT VARIABLES FOR FRONT LIFTING

Integration Level	External Loading	Trial	
I: (DHM only)	0 lb	1	2
	1 lb	1	2
	20 lbs	1	2
II: (DHM +MOCAP + mockup)	0 lb	1	2
	1 lb	1	2
	20 lbs	1	2
III: (DHM + MOCAP + VE)	0 lb	1	2
	1 lb	1	2
	20 lbs	1	2

4.4.2 Side Lifting Task, Pushing Task, Reaching Task

For the side lifting, pushing, and reaching task, the independent variable is Integration Level (3 levels). All subjects perform 2 trials of side lifting, pushing and reaching tasks in both Integration Level II and III, totally 12 trails. Researchers perform 2 trial of manual manipulation in Integration Level I for side lifting, pushing and reaching tasks, totally 6 trials. Following table shows the design of experiment for the side lifting, pushing and reaching tasks.

Table 5. INDEPENDENT VARIABLES FOR ALL TASKS

Integration Level (IL)	Trial
I: (DHM Simulation)	1
	2
II: (DHM + MOCAP +Mockup)	1
	2
III: (DHM + MOCAP + VE)	1
	2

4.5 Dependent Variables and Ergonomic Assessments

For each task, corresponding ergonomic assessment is performed to evaluate the injury risks and the ergonomic assessment results are the dependent variables. The 1991 revised NIOSH Lifting Equation (1991 NLE) and the Static Strength Prediction (SSP) are chosen to analyze lifting tasks. Rapid Upper Limb Assessment (RULA) is chosen to analyze the reaching task. The Static Strength Prediction (SSP) is chosen to analyze the pushing task. Following is the details for these ergonomic assessments.

4.5.1 Lifting Task

The 1991 revised NIOSH Lifting Equation is chosen as the ergonomic assessment for evaluating the injury risk from the lifting task. The 1991 NIOSH Lifting Equation takes the original and destination position and some general descriptive information as inputs. The parameters of the 1991 NIOSH Lifting Equation, such as the work duration, lifting frequency and handgrip coupling are kept unchanged for all experiment trials. Following table is the pre-setup value of those 1991 NIOSH Lifting Equation parameters.

Table 6. PARAMETERS OF NIOSH LIFTING EQUATION

Work hour	2 hours
Lift Frequency	2 lifts /min
Coupling	Good

1991 NIOSH Lifting Equation can be used to analyze both single and multiple lifting tasks (NIOSH, 1991; Waters, 1993, 1994). In this study, all lifting tasks are single lifting task. The output of 1991 NIOSH Lifting Equation, the Recommended Weight Limit (RWL) is used to predict the maximum loading weight to keep this lifting task safe. The RWL is the dependent variable of lifting tasks.

Besides the 1991 NIOSH Lifting Equation, the Static Strength Predication (SSP) is also used to assess the injury risk of lifting task. The SSP takes the body posture and external loading as inputs and report many things regarding human internal loading based on current external load and posture (Chaffin, 1984). The SSP assessment result, the trunk flexion/extension torque moment (Nm), is chosen as dependent variable of lifting task. For side lifting task, trunk rotation torque moment (Nm) is added as an extra dependent variable. Following table summarizes the dependent variables for lifting tasks.

Table 7. DEPENDENT VARIABLES OF LIFTING

	1991 NIOSH Lifting Equation	Static Strength Prediction
Front Lifting Task	RWL (lb)	Trunk Flex./Ext. Torque (Nm)
Side Lifting Task	RWL (lb)	Trunk Flex./Ext. Torque (Nm)
		Trunk Rotation Torque (Nm)

4.5.2 Reaching Task

For reaching task, the Rapid Upper Limb Assessment (RULA) is taken as the ergonomic assessment. The RULA takes the posture of upper arm as input and report a 1-7 ranked number to represent the risk injury of the upper limb activity (Corlett, 1999). The RULA 1-7 score is the dependent variable for reaching task. Corresponding parameters, other than posture information, are set constant for all reaching tasks. Following table is the pre-setup value of those RULA parameters.

Table 8. PARAMETERS OF RAPID UPPER LIMB ASSESSMENT

Body Group A	Muscle Use	Normal, no extreme use
	Force and Loads	<2kg intermittent load
	Arm Support	None
Legs and Feet		Standing, Weight even, Room for weight changes
Body Group B	Muscle Use	Normal, No extreme use
	Force and Loads	<2kg intermittent load

4.5.3 Pushing Task

For pushing task, the Static Strength Predication (SSP) is chosen as the ergonomic assessment. The shoulder torque moment (Nm) and elbow torque moment (Nm) are chosen to be dependent variables to evaluate the injury risk from pushing task (Samuelsson, et al.; 2004, Frisiello, et al.; 1994, Laursen, et al., 2002). The horizontal pushing force is set to 2.25 kg, based on our force gauge measurement on the initial force to push the tool cart.

4.6 Hypotheses

This study focuses on validating the performance of Virtual Build in ergonomic assessments. Following hypotheses are tested in this study.

1. Hypothesis one: the mean 1991 NIOSH lifting equation RWL value of Virtual Build with Integration Level I, II, and III is equal to criteria value.
   H0: $\mu_I = \mu_{II} = \mu_{III} = \text{criteria value}$
   H1: H0 is false

2. Hypothesis two: for all ergonomic assessments, the test-retest reliability of Virtual Build, the correlation of ergonomic assessment result between two trials, is in "Excellent" range (0.80, 1.0].

3. Hypothesis three: The Integration Level of Virtual Build has no significant effect on ergonomic assessments.
   H0: $\mu_I = \mu_{II} = \mu_{III}$
   H1: H0 is false

4.7 Design of Experiment

This study includes three experiment conditions that correspond to the three Virtual Build Integration Levels. Human subject will participate in experiment condition 2 and 3. Researchers will finish the experiment condition 1 without human subject participation.

4.7.1 Experiment Condition 1

The experiment condition 1 focuses on establishing a benchmark of using DHM in ergonomic research. The experiment condition one evaluates the reliability and validity of Virtual Build Integration Level I, which involves using DHM manually for ergonomic study. No human subject participates in experiment condition 1. Researchers use Inverse-Kinematics model to manipulate the digital manikin in UGS JACK software to finish task. The experiment condition 1 includes 6 tasks. They are

- Tasks1/2/3 : Front lifting (0 lb / 1 lb / 20 lbs)
- Task4 : Side lifting (l lb)
- Task5 : Forward reaching
- Task6 : Horizontal pushing

4.7.2 Experiment Condition 2

The experiment condition 2 focuses on evaluating the reliability and validity of Virtual Build Integration Level II, which integrates DHM and Motion Capture, with physical mockup. A physical mockup is built to simulate a real workstation. Human subjects are invited to perform tasks by interacting with the physical mockup.

The experiment condition 2 includes 6 tasks. These tasks are same as those of experiment condition 1.

- Tasks1/2/3 : Front lifting of 0 lb / 1 lb / 20 lbs
- Task4 : Side lifting of 1 lb
- Task5 : Forward reaching
- Task6 : A horizontal pushing

For each task in the experiment condition 2, human subject performs two trials.

4.7.3 Experiment Condition 3

The experiment condition 3 focuses on evaluating the reliability and validity of Virtual Build Integration Level III, which integrates DHM, Motion Capture and VE. Different from experiment condition 2, the CAD drawing file of the workstation will be imported into UGS JACK to generate the Virtual Environment, and no physical mockup is used in experiment condition 3. The workstation in the CAD drawing is exactly the same size as the physical mockup in experiment condition 2. Before experiment condition 3 starts, researchers moved the physical mockup outside the experiment volume. Human subject wore the motion capture suit and Head Mounted Display, and performed those tasks through interacting with the Virtual Environment. The experiment condition 3 includes the same 6 tasks as the experiment 1 does. They are

- Tasks1/2/3 : Front lifting (0 lb / 1 lb / 20 lbs)
- Tasks4 : Side lifting (l lb)
- Tasks5 : Forward reaching
- Tasks6 : A horizontal pushing/ pulling

All tasks are implemented collision detection to give subject additional feedback. During the lifting task, if the box is put on the desk or placed on the shelf, the border line of the box and desk/shelf will change to yellow color.

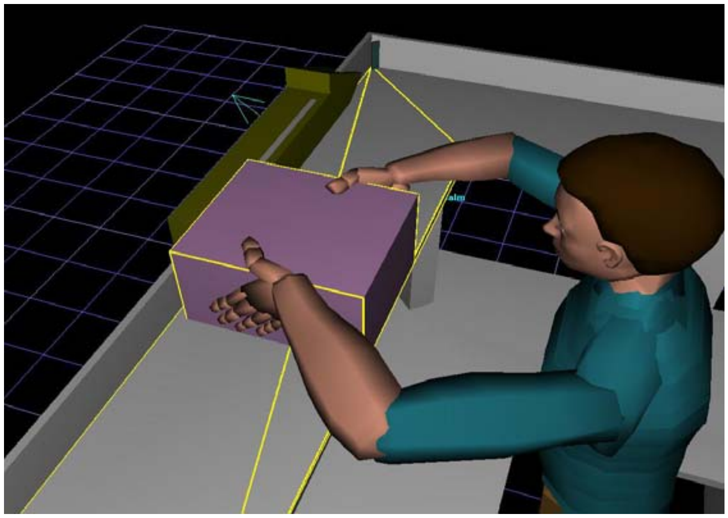

Figure 13 COLLISION DETECTION IN LIFTING

For the reaching task, if finger reaches the corner of shelf, both the finger and corner line will change to yellow color.

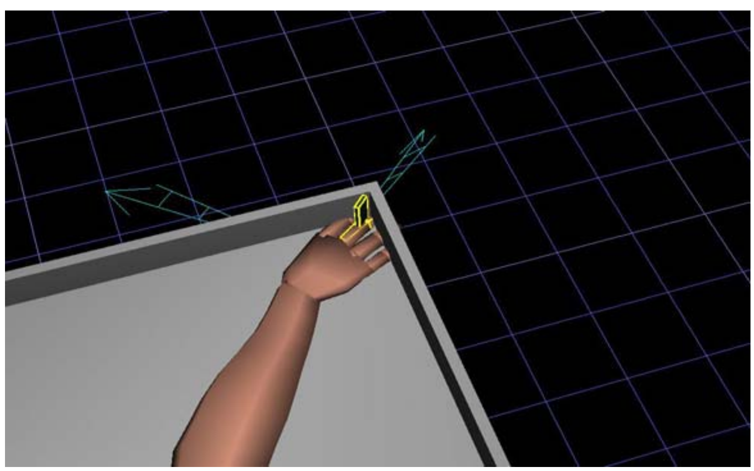

Figure 14 COLLISION DETECTION IN REACHING

For the pushing task the, the collision between subject palm and the handler of the tool cart is detected. If palms touch tool cart handler, both the palms and tool cart handler will change to yellow color.

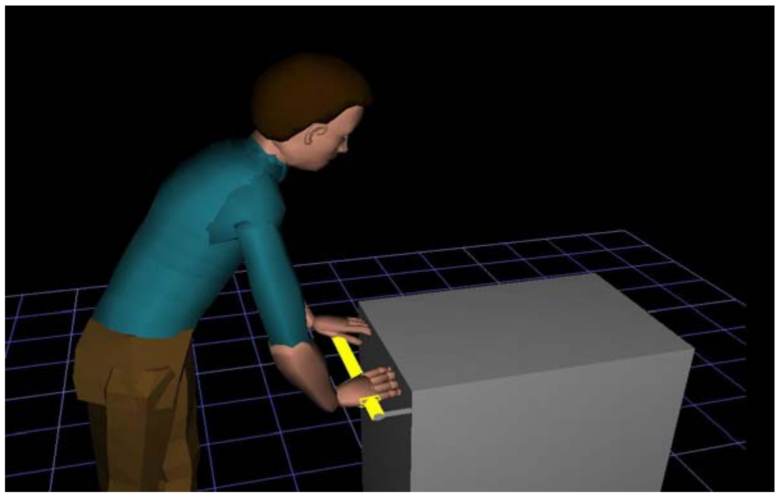

Figure 15 COLLISION DETECTION IN PUSHING

4.8 Procedure

Human subject was invited to Human System Engineering lab at the Center for Advanced Vehicular System in Mississippi State University. Before the experiment, subject read the Informed Consent Form and signed on it if he/she agreed terms and conditions. The Informed Consent Form had been approved by Mississippi State University IRB committee before subject recruitment (Please refer appendix B for the Informed Consent Form). Then subject filled the demography and musculoskeletal injury history forms. After that, researchers helped the subject wear the motion capture suit and placed 34 reflective markets on it at landmark joint or locations (Please refer Appendix A for the motion capture suit markers location). Then researchers took anthropometric measurements, the standing height and weight, of the subject, following procedure that is defined by Kroemer (1999) and McArdle et al. (2001). All standing height are with shoes height.

The motion capture system was well-calibrated before the subject entering the working volumes. After the experiment introduction, subject was shown to the working volume, and given the instruction to finish tasks. The order of starting with either Integration Level II or III was randomized across subjects. The order of lifting, reaching and pushing was fully randomized across subjects. However, due to experiment setup reason, all lifting tasks are grouped together to facilitate the setup. That means, if one subject started with lifting task, he/she finished all lifting tasks first, and then continued to either pushing or reaching task. And within the lifting group, the order of different External Loading weights for front lifting and side lifting was fully randomized.

4.9 Data Analyses

This study intends to evaluate the validity and reliability of Virtual Build methodology. Corresponding analyses are performed to evaluate the validity and reliability.

4.9.1 Reliability Analysis

The correlation of ergonomic results between two trials of each task is used to evaluate the test-retest reliability, which shows the over-time consistence. The *ICC* will be calculated, and then interpreted according to the Table 3 to evaluate the over-time reliability.

4.9.2 Validity Analysis

Validity is evaluated by comparing the 1991 NIOSH Lifting Equation RWL result with the criteria result. Manual measurement and calculation of 1991 NIOSH Lifting Equation RWL is treated as the error-free criteria. The statistical test is used to find out is there any significant difference between the Virtual Build NIOSH RWL result and the criteria RWL result. Corresponding statistics are used to verify whether the difference between experiment results and criteria result is negligible.

The comparison of the ergonomic assessments result among all three Virtual Build Integration levels is used to evaluate whether Integration level has a significant effect on the Virtual Build's performance. The statistical test is used to find out is there any significant difference between ergonomic assessments results among the three Virtual Build Integration levels.

CHAPTER V

RESULTS

5.1 Subject

36 subjects, 23 male and 13 female, participated in this study. Age ranges from 19 to 48. The overall average standing height is 173.5cm. The mean standing heights are 178.6 cm and 164.3 cm for male and female subject respectively. The average values are very close to the 50th percentile standing height of North American population, 179cm and 165cm for male and female respectively. The range of standing height is 152cm to 192cm, which covers from 5th percentile female (154cm) to 95th percentile male (190 cm) The following picture shows using the normal distribution to fit the overall, male and female standing height.

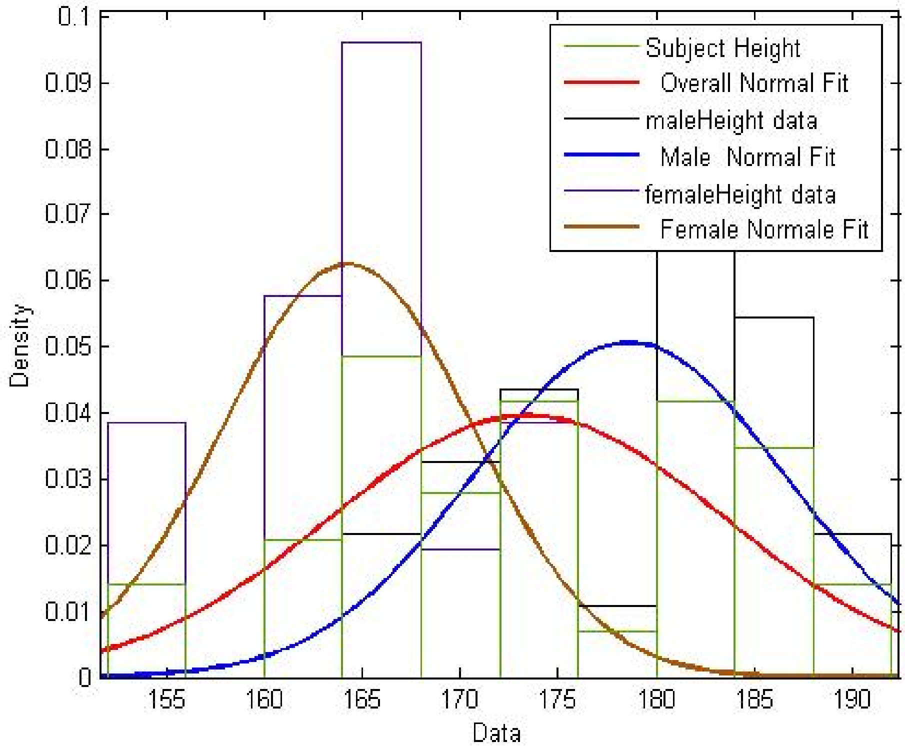

Figure 16 HISTOGRAM AND NORMAL FIT OF STANDING HEIGHT

The following figure is the normal probability plot of overall standing height.

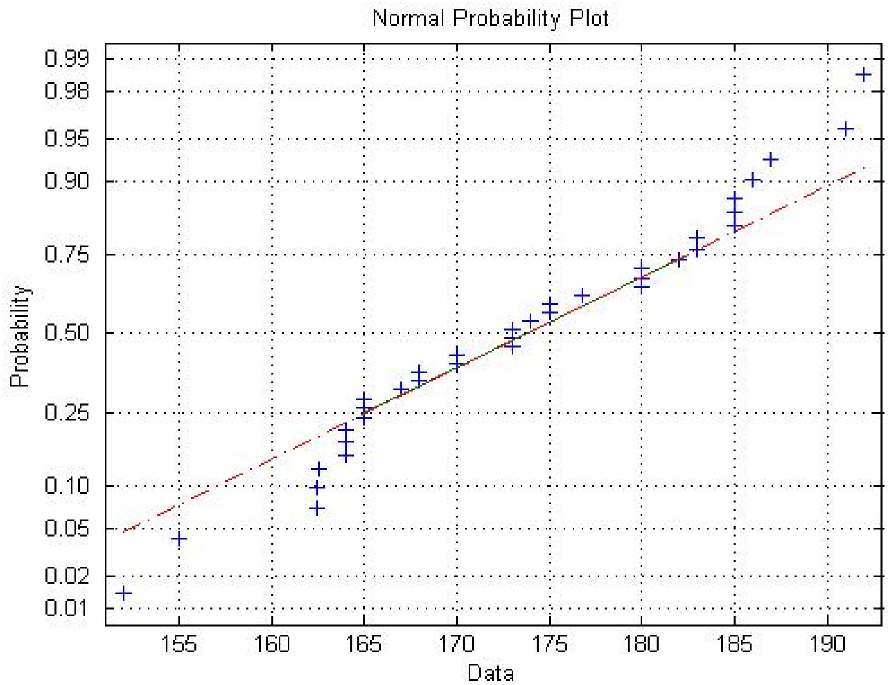

Figure 17 NORMAL PROBABILITY PLOT OF STANDING HEIGHT

Figure 17 tells that the standing height is close to the solid line that fit from the 1 quartile to 3 quartile distribution, which is the linear fit of the normal distribution. The Kolmogorov-Smirnov goodness-of-fit test result is d=0.103894, and the probability of $Pr > d$ is bigger than 0.1500, shows that the overall subject standing height has a normal distribution.

5.2 Statistics

Figure 18 is the plot of the Virtual Build 1991 NIOSH Lifting Equation RWL result of all four lifting tasks in all three Virtual Build Integration Levels. Each small plot figure is for one lifting task, and three lines in each plot figure shows the Integration Level I, II and III respectively.

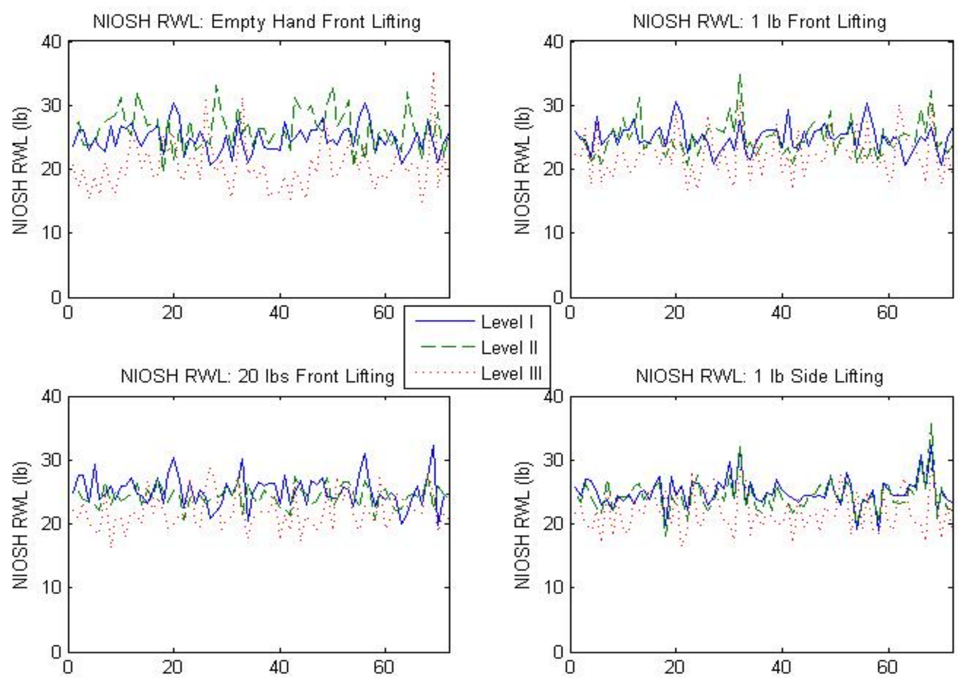

Figure 18 PLOT OF NIOSH LIFTING EQUATION RWL

Figure 18 tells that the dot line, which represents the Virtual Build Integration Level III, has bigger wave than other two lines. The following table summarizes the mean value and standard deviation of the NIOSH RWL of each lifting task in each Virtual Build Integration Level.

Table 9. SUMMARY STATISTICS OF NIOSH LIFTING EQUATION RWL

Integration Level	0 lb Front Lifting	1 lb Front Lifting	20 lbs Front Lifting	1 lb Side Lifting
I: DHM Simulation	$\mu = 24.79$ $\sigma = 2.1677$	$\mu = 25.069$ $\sigma = 2.1749$	$\mu = 25.165$ $\sigma = 2.4145$	$\mu = 25.139$ $\sigma = 2.2329$
II:	$\mu = 26.298$	$\mu = 24.85$	$\mu = 24.314$	$\mu = 24.429$

DHM + MOCAP + mockup	$\sigma = 3.037$	$\sigma = 2.621$	$\sigma = 1.99$	$\sigma = 2.594$
III: DHM + MOCAP + VE	$\mu = 20.483$ $\sigma = 3.814$	$\mu = 22.203$ $\sigma = 3.423$	$\mu = 21.932$ $\sigma = 2.905$	$\mu = 21.789$ $\sigma = 3.195$

A very interesting finding is that the UGS JACK NIOSH Lifting Equation analysis tool does not identify the side lifting task. The Virtual Build reports the Asymmetrical Multiplier 1, which means 0°of twisting, for the side lifting task. The actual side lifting task has a 90°twisting. This error may cause by a bug of UGS JACK. The NIOSH Lifting Equation RWLs in the upper table are results of the Virtual Build. The actual result of side lifting should be this result times 0.71, which is the Asymmetrical Multiplier of 90°. For the rest part of this study, all NIOSH RWL of side lifting is based on Virtual Build result, instead of the actual result, which is 0.71 times of the Virtual Build result. From the Table 10, we can tell that the variance of the NIOSH RWL in Integration Level III is bigger than that in either Integration Level I or II. This result is also shown in the Figure 18.

In each Virtual Build Integration Level, the variance of the 1991 NIOSH Lifting Equation RWL of empty hand front lifting is bigger than the variance of the 1991 NIOSH Lifting Equation RWL of lifting with box in hand, either 1 lb or 20 lbs. Without box in the hand, human subject performs the lifting task only based on his/her perceptual judgment; the result may reveal that the necessary external feedback from the box in hand will lower the variance in the human subject performance.

5.3 Reliability

The correlation of ergonomic results between two trials of each task is used to evaluate the test-retest reliability, which shows the consistency over time. The Intra Class Correlation (*ICC*) is calculated and interpreted according to the Table 3 in chapter 2 to evaluate the over-time reliability.

5.3.1 NIOSH Lifting Equation

The Intra Class Correlation of the 1991 NIOSH Lifting Equation RWL result between two trials is calculated and shown in following table. A cell with ICC value less than 0.8 is placed an asteroid (*).

Table 10. *ICC* OF NIOSH LIFTING EQUATION RWL

Integration Level	0 lb Front Lifting	1 lb Front Lifting	20 lbs Front Lifting	1 lb Side Lifting
I: DHM Simulation	*ICC* = 0.8664	*ICC* = 0.92906	*ICC* = 0.92498	*ICC* = 0.85674
II: DHM+MOCAP+ Mockup	*ICC* = 0.8342	*ICC* = 0.87931	*ICC* = 0.81592	*ICC* = 0.86894
III: DHM+MOCAP+ VE	*ICC* = 0.75889*	*ICC* = 0.79898*	*ICC* = 0.81588	*ICC* = 0.82298

All three Integration Level have *ICC* score on 1991 NIOSH Lifting Equation RWL value bigger than 0.8 for all four lifting tasks, except the 0 lb and 1 lb front lifting task in Integration Level III. The result shows that the Virtual Build Integration Level I and II provide ‘Excellent’ test-retest reliability on the 1991 NIOSH Lifting Equation analysis. The *ICC* score of the Integration Level III are smaller than those of Integration Level I and II. The test-retest reliability of the Integration Level III for empty hand front lifting and 1 lb front lifting are only ‘Good’.

Following is the plot of the correlation of the 1991 NIOSH Lifting Equation RWL value between two trials. The x and y axis values are the 1991 NIOSH Lifting Equation RWL value of trial one and two respectively. The square marker and cross marker denote the Integration Level I result and Integration Level II result respectively. They scatter very close to the straight line, shows a very good correlation. The circle marker denotes the Integration Level III result.

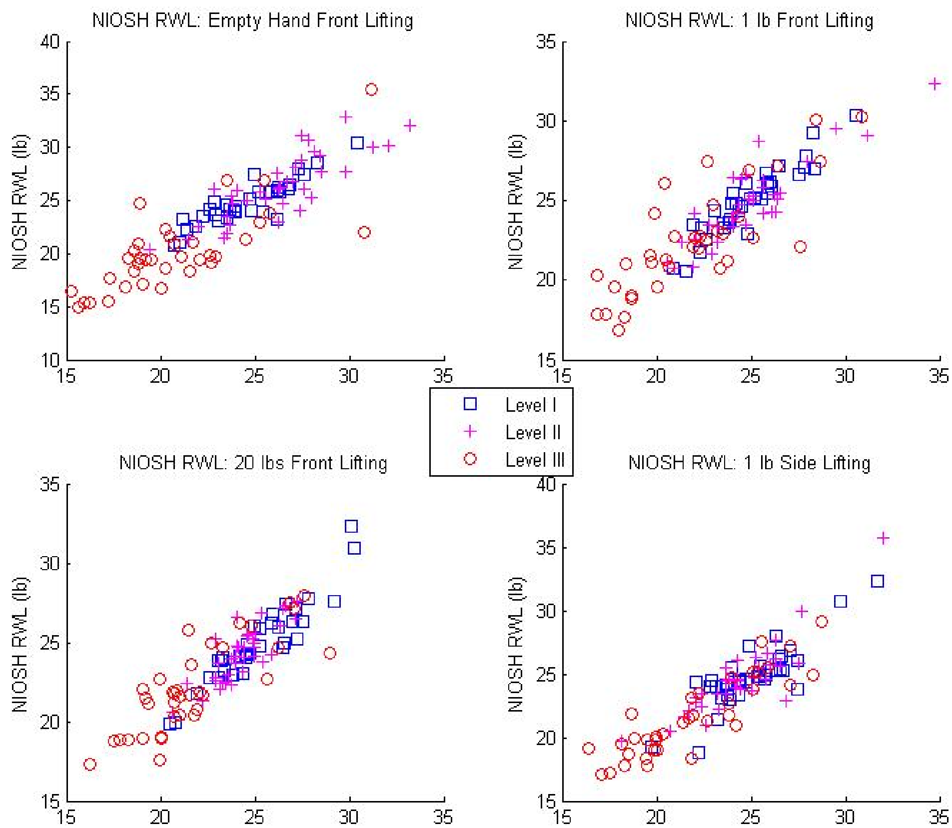

Figure 19 PLOT OF NIOSH LIFTING EQUATION RWL

5.3.2 Static Strength Prediction

The Static Strength Prediction (SSP) assessment relies on the human's posture. The Intra-Class Correlation of the SSP result between two trials of each subject can provide information of the over-time reliability of human subject's posture to finish the task. All lifting tasks and pushing task are analyzed using Static Strength Prediction.

5.3.2.1 Lifting Task

Both the original and destination position of each lifting task are analyzed using Static Strength Prediction assessments. The trunk flexion torque is chosen to evaluate the risk from the lifting. For the side lifting task the trunk rotation torque is also analyzed. The Intra-Class Correlation of the trunk torque value at the original position between two trials is listed in following table. The ICC value less than 0.8 is placed an asteroid at that cell.

Table 11. *ICC* OF TRUNK TORQUE: AT ORIGINAL POSITION OF LIFTING

Integration Level	0 lb Front Lifting	1 lb Front Lifting	20 lbs Front Lifting	1 lb Side Lifting	
	Trunk Flex./Ext.	Trunk Flex./Ext.	Trunk Flex./Ext.	Trunk Flex./Ext.	Trunk Rotation
I: DHM Simulation	*ICC* = 0.89882	*ICC* = 0.98590	*ICC* = 0.96727	*ICC* =0.90154	*ICC* =0.87602
II: DHM + MOCAP + Mockup	*ICC* = 0.87898	*ICC* = 0.92070	*ICC* = 0.82433	*ICC* = 0.91963	*ICC* = 0.83583
III: DHM + MOCAP + VE	*ICC* = 0.95042	*ICC* = 0.90612	*ICC* = 0.94501	*ICC* = 0.9664	*ICC* = 0.77618*

All three Integration Levels have *ICC* scores on trunk torque value bigger than 0.8 for all four lifting tasks, except the trunk rotation torque value in Integration Level III. The results show that the Virtual Build Integration Level I and II, and III, provide 'excellent' test-retest reliability on the Static Strength Prediction assessment at the original position of lifting tasks. The Intra Class Correlations of the Integration Level III are smaller than those of Integration Level I and II. The test-retest reliability of the Integration Level III for trunk rotation torque is only 'good'. Figure 20 is the plot of the correlation of the Static Strength Prediction assessment results between two trials. The x and y axis value are the Static Strength Prediction value of trial one and two respectively. The square marker and cross marker and circle marker denote the Integration Level I, II and III result respectively.

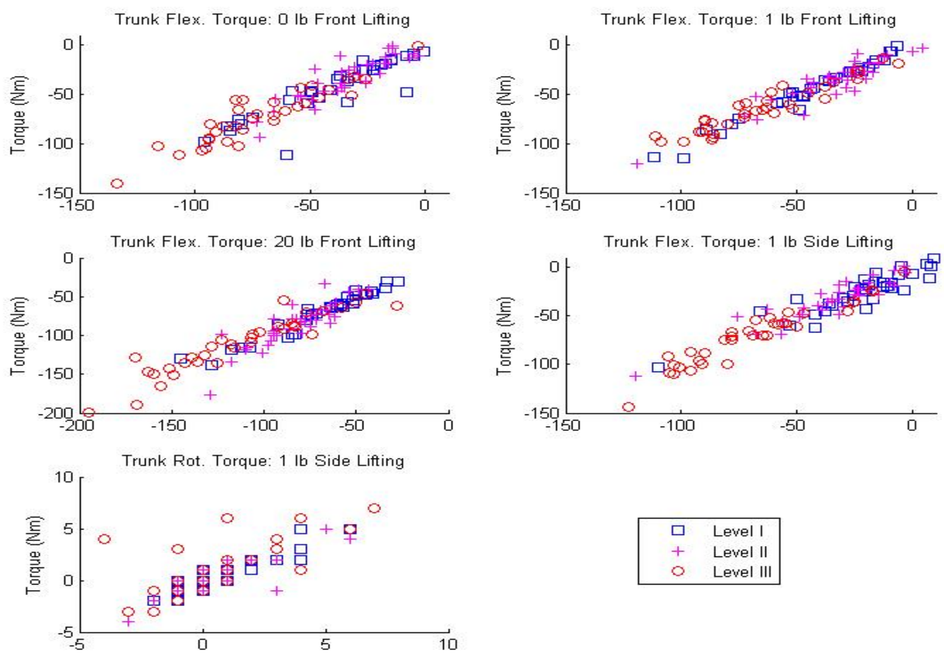

Figure 20 PLOT OF TRUNK TORQUE AT ORIGINAL POSITION: LIFTING

The Intra-Class Correlation of the trunk torque value at the destination position between two trials is calculated in following table. The ICC value less than 0.8 is placed an asteroid at that cell.

Table 12. *ICC* OF TRUNK TORQUE : AT DESTINATION POSITION OF LIFTING

Integration Level	0 lb Front Lifting	1 lb Front Lifting	20 lbs Front Lifting	1 lb Side Lifting	
	Trunk Flex./Ext.	Trunk Flex./Ext.	Trunk Flex./Ext.	Trunk Flex./Ext.	Trunk Rotation
I: DHM Simulation	*ICC* = 0.96675	*ICC* = 0.96939	*ICC* = 0.97528	*ICC* = 0.98141	*ICC* = 0.96255
II: DHM + MOCAP + Mockup	*ICC* = 0.94908	*ICC* = 0.95339	*ICC* = 0.85788	*ICC* = 0.95171	*ICC* = 0.95175
III: DHM + MOCAP + VE	*ICC* = 0.90478	*ICC* = 0.84571	*ICC* = 0.86412	*ICC* = 0.90928	*ICC* = 0.86010

All three Integration Levels have Intra-Class Correlation on Static Strength Prediction trunk torque value bigger than 0.8 for all four lifting tasks. The results show that the Virtual Build Integration Level I and II, as well as III, provide 'excellent' test-retest reliability on the Static Strength Prediction at the destination position of lifting tasks.

Figure 21 is the plot of the correlation of the Static Strength Prediction assessment results at the destination position of lifting task between two trials. The x axis value is the trial one Static Strength Prediction value and y axis is the trial two values. The square marker and cross marker and circle marker denote the Integration Level I, II and III result respectively. They scatter very close to the straight line, which shows a perfect correlation between two trials.

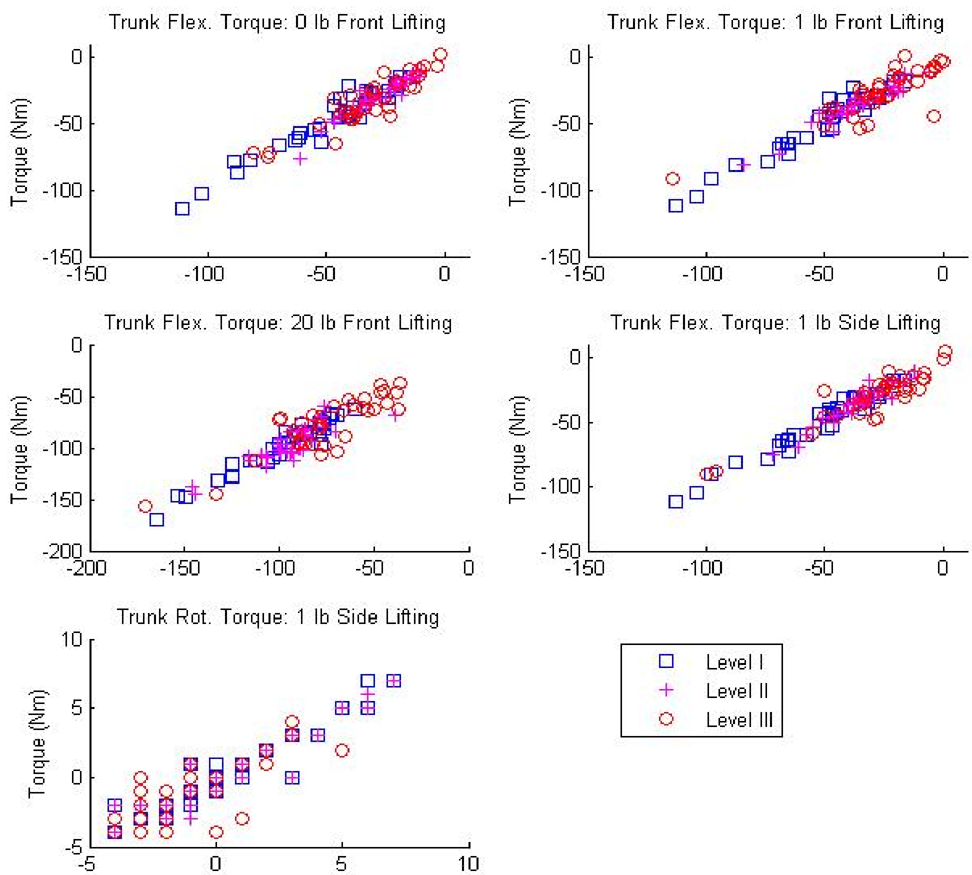

Figure 21 PLOT OF TRUNK TORQUE AT DESTINATION POSITION: LIFTING

5.3.2.2 Pushing Task

The shoulder abduction torque and elbow torque are chosen to evaluate the pushing task. The Intra Class Correlation of the shoulder torque and the elbow torque value between two trials are calculated and listed in following table. The ICC value less than 0.8 is placed an asteroid (*) at that cell.

Table 13. *ICC* OF SHOULDER, ELBOW TORQUE: PUSHING

Integration Level	Shoulder (Abduction)	Elbow
I: DHM Simulation	*ICC* = 0.91474	*ICC* = 0.97010
II: DHM + MOCAP + Mockup	*ICC* = 0.90001	*ICC* = 0.86176
III: DHM + MOCAP + VE	*ICC* = 0.91947	*ICC* = 0.9314

All three Integration Levels have Intra-Class Correlation on Static Strength Prediction trunk torque value bigger than 0.8. The results show that the Virtual Build Integration Level I and II, as well as III, provide ‘excellent’ test-retest reliability on the Static Strength Prediction for the pushing task analysis. Figure 22 is the correlation plot of the Static Strength Prediction assessment results of pushing task between two trials.

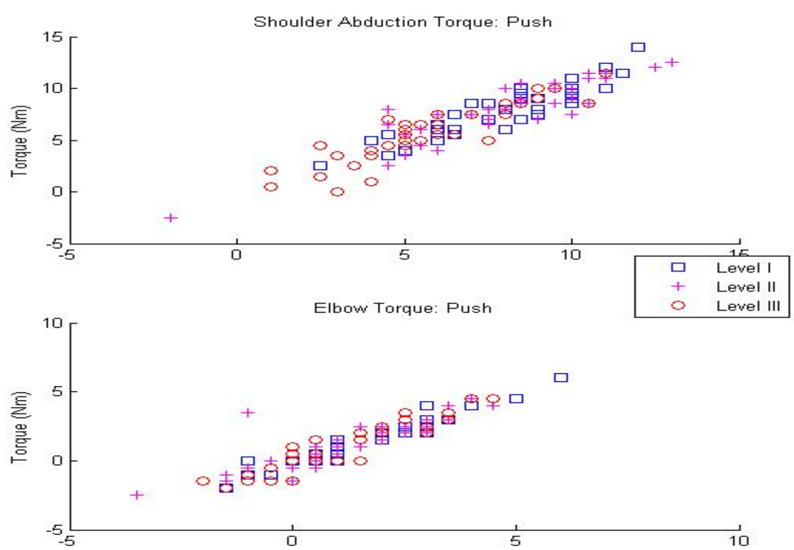

Figure 22 PLOT OF SHOULDER, ELBOW TORQUE: PUSHING

The square marker and cross marker and circle marker denote the Integration Level I, II and III result respectively. They scatter very close to the straight line, which shows a perfect correlation between two trials.

5.3.3 Rapid Upper Limb Assessment

RULA scores an integer ranking from 1 to 7. The Intra Class Correlation and Pearson's r correlation are for continuous variable. Both of them can not be used to evaluate the test-retest reliability of RULA score. Spearman correlation is used for RULA result. Spearman rank correlation is a distribution-free analog of correlation analysis. It can be applied to compare two independent random variables, each at several levels. Spearman's rank correlation works on ranked data. The Spearman's r_s coefficient indicates agreement. A value of r_s near one indicates good agreement; a value near zero, poor agreement. As a distribution-free method, the Spearman rank correlation does not make any assumptions about the distribution of the underlying data.

Table 14. SPEARMAN'S R OF RULA: REACHING

Integration Level	I: DHM	II: DHM+MOCAP+Mockup	III: DHM+MOCAP+VE
Spearman's r_s	1.0 $p<0.0001$	1.0 $p<0.0001$	1.0 $P<0.0001$

The spearman's r correlation of RULA score between two trials shows perfect match for all three Integration Levels.

5.4 Validity

5.4.1 Criteria

Validity is evaluated by comparing the 1991 NIOSH Lifting Equation RWL result of the Virtual Build with the criteria result. Manually measured and calculated of 1991 NIOSH Lifting Equation RWL is treated as the error-free criteria. Four trials of manual measurements were done and hand calculation was performed. The 1991 NIOSH Lifting Equation frequency parameter is 2 lifting per minute, and work duration is 2 hours, 6 minutes break, and hand coupling is good. Following the table is the measurement of each trial and the corresponding 1991 NIOSH Lifting Equation RWL. The average value of the 1991 NIOSH Lifting Equation RWL from the 4 trials, 24.74 lb is the criteria.

Table 15. MANUAL CALCULATION OF NIOSH LIFTING EQUATION RWL

Trial	H(cm)	V(cm)	D(cm)	NIOSH RWL(lbs)
1	34.9	80.9	60	24.74
2	34	81	60	25.35
3	35.3	80.9	60.2	24.45
4	35.4	81	60.1	24.4

5.4.2 Root Mean Square Error (RMSE)

The mean squared error (MSE) is used to evaluate the deviation of a measurement from the target. The square root of MSE, RMSE, which has the same unit as the measurement, is often chosen to substitute the MSE to measure the error. Following table is RMSE of the 1991 NIOSH Lifting Equation RWL of each lifting task in all Virtual Build Integration Levels.

Table 16. RMSE OF NIOSH LIFTING EQUATION RWL

Integration Level	0 lb Front Lifting	1 lb Front Lifting	20 lbs Front Lifting	1 lb Side Lifting
I: DHM Simulation	*RMSE* =2.15	*RMSE* =2.18	*RMSE* =2.44	*RMSE* = 2.25
II: DHM + MOCAP + Mockup	*RMSE* =3.39	*RMSE* =2.61	*RMSE* =1.94	*MSE* =2.59
III: DHM + MOCAP + VE	*RMSE* =5.698	*RMSE* =4.241	*RMSE* =4.03	*MSE* =4.33

The table 16 shows that the Integration Level III has bigger RMSE than Integration Level II and Level II, for all lifting tasks. In both Integration Level II and III, the empty hand lifting task has the biggest RMSE among the all lifting tasks. RMSEs of the four lifting tasks in Integration Level I are quite close to each other. Following figure is the bar chart of RMSE of each lifting task in all three Integration Levels. The Integration Level III has higher RMSE bar than both Integration Level I and II.

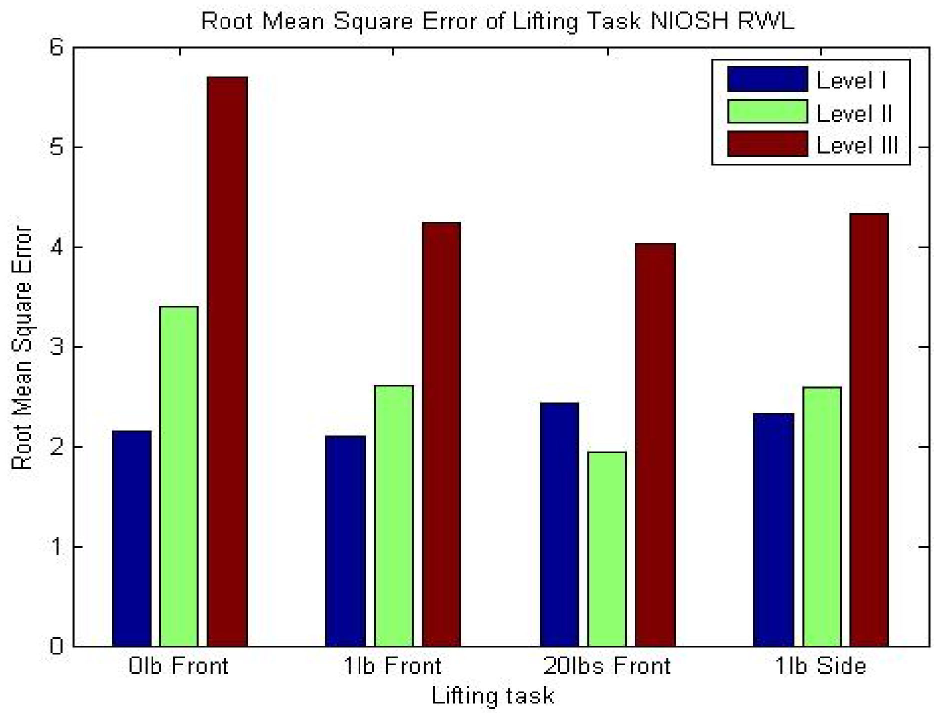

Figure 23 BAR CHART OF RMSE OF NIOSH LIFTING EQUATION RWL

5.4.3 *T* Test

T test was run to test whether the 1991 NIOSH Lifting Equation RWL of each task in each environment is equal to the criteria. The sample size (72) is big enough to assume the result is normal distributed. The hypothesis is

H0: $\mu_{ij} = 24.74$ *i = 1 – 3 : 3 Integration Levels*
j = 1 – 4 : 4 lifting tasks

H1: H0 is false

Following table is the results from *t* test:

Table 17. *T*-TEST: NIOSH LIFTING EQUATION RWL

Integration Level	0 lb Front Lifting	1 lb Front Lifting	20 lbs Front Lifting	1 lb Side Lifting
I: DHM Simulation	t = 0.2* p =0.8441 df = 71	t = 1.28* p =0.2036 df = 71	t = 1.49* p =0.1397 df = 71	t = 1.52* p =0.1337 df = 71
II: DHM + MOCAP + mockup	t = 4.35 $p < 0.0001$ df = 71	t = 0.35* p =0.7238 df = 71	t = -1.90* p =0.0755 df = 71	t = -1.02* p =0.3118 df = 71
III: DHM + MOCAP + VE	t = -9.47 $p < 0.0001$ df = 71	t = -6.29 $p < 0.0001$ df = 71	t = -8.20 $p < 0.0001$ df = 71	t = -7.84 $p < 0.0001$ df = 71

In the table 17, cells with '*' are with *p* value bigger than 0.05. We can not reject *H0* in those cells.

From the result in Table 18, we can tell that:

1. Virtual Build Integration Level I can provide a valid 1991 NIOSH Lifting Equation assessment for all lifting tasks.

2. Virtual Build Integration Level II can not provide a valid 1991 NIOSH Lifting Equation assessment for the empty hand front lifting task. The mean value is not the criteria 1991 NIOSH Lifting Equation RWL value at the 0.05 level of significance. But Virtual Build Integration Level II provide a valid 1991 NIOSH Lifting Equation assessment for lifting tasks with box, either 1 lb or 20 lbs.

3. For all lifting tasks in Integration Level III, we can not accept the null hypothesis, as all *p* value is less than 0.05. All 1991 NIOSH Lifting Equation RWL results are significantly different from criteria result. Lifting with box or without box in hand, the mean value of 1991 NIOSH Lifting Equation RWL values are not the criteria value at the 0.05 significant level.

5.4.4 ANOVA Test

5.4.4.1 NIOSH Lifting Equation RWL

There are two factors for the front lifting task. The factors are Integration Level (3 levels) and the External Loading (3 levels). For the 1991 NIOSH Lifting Equation RWL score, the hypotheses that are tested are as follows:

1. *H0:* $\mu_I = \mu_{II} = \mu_{III}$ (no effect of Integration Level)
   H1: *H0* is false

2. *H0:* $\mu_0 = \mu_1 = \mu_{20}$ (no effect of External Loading)
   H1: *H0* is false

3. *H0:* $\mu_{I0} = \mu_{I1} = \ldots = \mu_{ij}$ (no effect of interaction of Integration Level and External Loading)
   i: I, II, III (Integration Level)
   j: 0, 1, 20 (External Loading)
   H1: *H0* is false

Two-way ANOVA analysis is performed to identify is there significant difference of the 1991 NIOSH Lifting Equation RWL result between the Integration Levels, External Loadings and their interactions. Fisher LSD post hoc analysis is conducted to compare the mean values of 1991 NIOSH Lifting Equation RWL among the combination of the two independent variables. Following tables are ANOVA results:

Table 18. ANOVA RESULT: PART ONE

Source	DF	SS	MS	F	P
Model	8	2073.3	259.1	34.15	<0.0001
Error	639	4849.2	7.6		
Corrected Total	647	6922.6			

Table 19. ANOVA RESULT: PART TWO

Source	DF	Type I SS	MS	F	P
Integration Level (IL)	2	1793.4	896.7	118.2	<0.0001
External Loading (EL)	2	4.9	2.45	0.32	0.7238
IL * EL	4	275.0	68.7	9.06	<0.0001

The ANOVA result shows ($F = 34.15$, $p<0.0001$, $df = 8$) the p value less than 0.05, we can not accept the null hypothesis that the mean value of 1991 NIOSH Lifting Equation RWL are equal. The ANOVA result of the source of the interaction of Integration Level and External Loading is $F = 9.06$, $p < 0.0001$, $df = 4$. The p value is less than 0.05. It shows that the interaction of Integration Level and External Loading has a significant effect on the 1991 NIOSH Lifting Equation RWL result. Since the interaction has significant effect, the main effects of the factors involved in the interaction may not have much practical interpretative value.

Following figure shows the effect of interaction of Integration Level and External Loading on the 1991 NIOSH Lifting Equation RWL result.

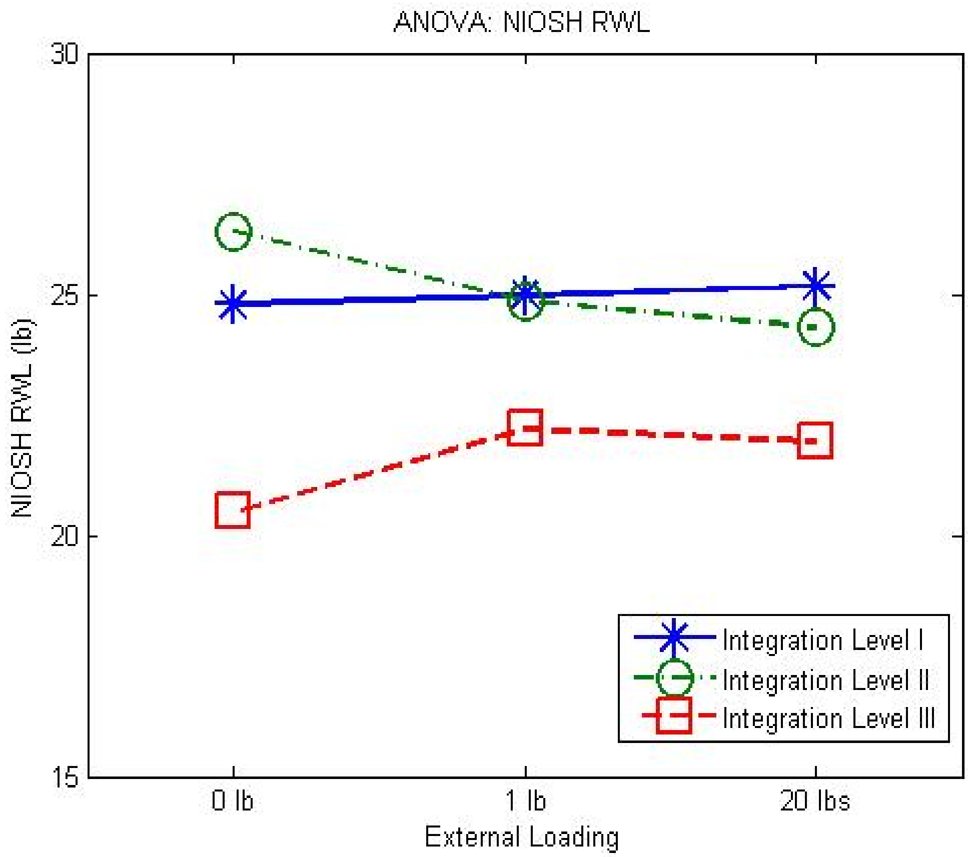

Figure 24 ANOVA OF NIOSH LIFTING EQUATION RWL

The line with star marker shows the 1991 NIOSH Lifting Equation RWL in Integration Level I. The External Loading (0/1/20 lb) has no significant effect on the assessment result. The line with circle and square marker shows the 1991 NIOSH Lifting Equation RWL in Virtual Build Integration Level II and II respectively.

Fisher Lest Significant Difference (LSD) analysis is conducted to compare the mean values of 1991 NIOSH Lifting Equation RWL among combinations of the two factors and also group the combinations with no significant difference. Following table is the result of Fisher LSD test.

Table 20. FISHER-LSD: NIOSH LIFTING EQUATION RWL

Group	1991 NIOSH Lifting Equation RWL Mean	Integration Level * External Loading
A	26.298	IL: II * EL: 0 lb
B	25.165	IL: I * EL: 20 lbs
	24.974	IL: I * EL: 1 lb
	24.850	IL II * EL: 1 lb
	24.790	IL: I * EL: 0 lb
	24.314	IL: II * EL: 20 lbs
C	22.204	IL: III * EL:1 lb
	21.932	IL: III * EL: 20 lb
D	20.483	IL: III * EL: 0 lb

From the Fisher LSD post hoc analysis, there are 4 groups, which are denoted by A, B, C, and D. It shows that:

1. The mean values of 1991 NIOSH Lifting Equation RWL of all three External Loadings in Integration Level I belongs to the same group, and there is no significant difference among them.

2. For the three External Loadings in Integration Level II, the mean 1991 NIOSH Lifting Equation RWL value of empty hand lifting (EL: 0 lb) belongs to different group from those of lifting with box in hand (EL: 1 lb or 20 lbs).

3. For the three External Loadings in Integration Level III, the mean 1991 NIOSH Lifting Equation RWL value of empty hand lifting (EL: 0 lb) belongs to different group from those of lifting with box in hand (EL: 1 lb or 20 lbs). The mean 1991 NIOSH Lifting Equation RWL values of all three External Loadings in the Integration Level III belong to different group from Integration Level I or II.

4. The 1991 NIOSH Lifting Equation RWL result of lifting task with box in Integration Level II belong to the same group of those of lifting tasks in Integration Level I. This group is the valid group.

For the side lifting task, the external loading is always1 lb. The one-way ANOVA with Fisher LSD is performed to find out the difference of 1991 NIOSH Lifting Equation RWL value among three Integration Levels. Following table is the ANOVA result.

Table 21. ANOVA: NIOSH LIFTING EQUATION RWL: SIDE LIFTING

$F = 27.52$ $p < 0.0001$ $df = 2$
Fisher LSD:
A: 24.9317 : IL: I
A : 24.4286: IL: II
B: 21.7889: IL: III

The ANOVA result ($F = 27.52$, $p < 0.0001$, $df = 2$) tells that there is significant difference in the mean 1991 NIOSH Lifting Equation RWL value among the three Integration Levels for the side lifting task. Fisher LSD test tells that the mean 1991 NIOSH Lifting Equation RWL value of Integration Level I and II belongs to the same group, which is different from that of Integration Level III.

5.4.4.2 Static Strength Prediction

Unlike the 1991 NIOSH Lifting Equation, which takes no information from the human subject posture, the Static Strength Prediction (SSP) assessment relies on the human posture and external loading. The SSP result: torque moment, can add more ergonomic information to analyze tasks.

5.4.4.2.1 Lifting Task

The trunk flexion/extension torque is chosen to analyze the injury risk from lifting task, and the score of the trunk flexion/extension torque is a dependent variable to evaluate the lifting task. For the side lifting task, the trunk rotation torque is added. Both the original position and destination position of each lifting task are analyzed using Static Strength Prediction. One-way ANOVA with Fisher LSD is used to identify the difference of SSP results among the three Integration Levels.

Original posture is at what human subject lift up the box from the platform of table, like Figure 8. Following table is the ANOVA result, with Fisher LSD, of the trunk torque at the original position of the lifting tasks

Table 22. ANOVA: TRUNK TORQUE: ORIGINAL POSITION

	0 lb Front Lifting	1 lb Front Lifting	20 lbs Front Lifting	1 lb Side Lifting	
	Trunk Flex./Ext.	Trunk Flex./Ext.	Trunk Flex./Ext.	Trunk Flex./Ext.	Trunk Rotation
ANOVA	*F = 50.24* *p <0.001* *df = 2*	*F*=28.93 *p <0.001* *df = 2*	*F*= 33.88 *p <0.0001* *df = 2*	*F*= 66.21 *p <0.0001* *df = 2*	*F*= 0.88 *p = 0.416* *df =2*
Fisher LSD	Grp. | Mean | IL A : -32.82 : II A : -39.75 : I B : -75.67 : III	Grp. | Mean | IL A : -35.28 : I A : -41.72 : II B : -68.5 : III	Grp. | Mean | IL A : -79.42 : I A : -84.24 : II B : -113.90 : III	Grp. | Mean | IL A : -23.90 : I A : -33.97 : II B : -72.31 : III	Grp. | Mean | IL A : 0.875 : I A : 0.528 : III A : 0.412 : II

One way ANOVA reports that the Integration Level has significant effects on the trunk flexion torque at the original lifting posture for all lifting task, but no effect on trunk rotation torque for side lifting task. For all lifting tasks at the original position, the absolute value of the mean trunk flexion/extension torque of Integration Level III is significant bigger than that in Integration Level I or II. For all front lifting tasks, the trunk flexion/extension torques at the original position in both Integration Level I and II belong to the same group. Figure 25 is the boxplot of the original position trunk torques of all lifting tasks in all three Integration Levels.

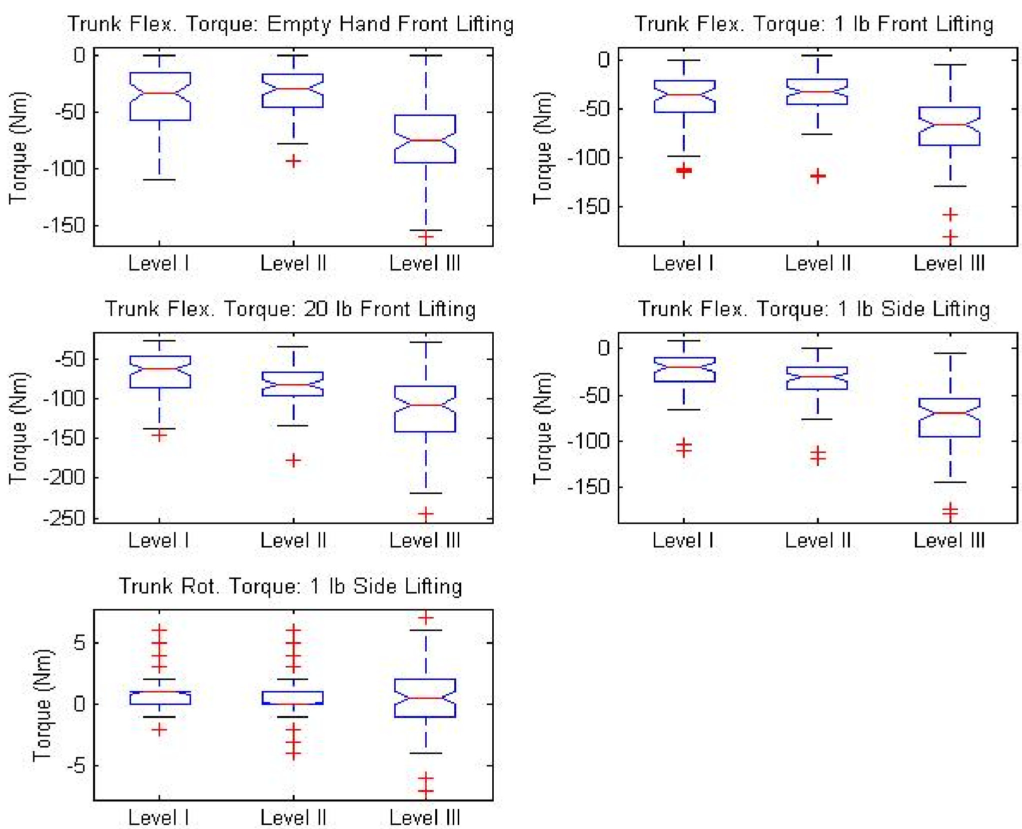

Figure 25 BOXPLOT OF TRUNK TORQUE: ORIGINAL POSITION

Destination posture is at what human subject place the box on the front upper shelf for front lifting task and right side shelf for side lifting task, which are shown in Figure 9 and 10. When the box is placed on the shelf and aligned with the shelf border line, the subject's posture is the destination posture. Following table is the ANOVA result, with Fisher LSD, of the trunk torque at the destination position of all lifting tasks

Table 23. ANOVA: TRUNK TORQUE: DESTINATION POSITION

	0 lb Front Lifting	1 lb Front Lifting	20 lbs Front Lifting	1 lb Side Lifting	
	Trunk Flex./Ext.	Trunk Flex./Ext.	Trunk Flex./Ext.	Trunk Flex./Ext.	Trunk Rotation
ANOVA	$F = 14.86$ $p < 0.001$ $df = 2$	F=25.91 $p < 0.001$ $df = 2$	F= 17.13 $p < 0.0001$ $df = 2$	F= 20.12 $p < 0.0001$ $df = 2$	F= 2.19 $p = 0.114$ $df = 2$
Fisher LSD	Grp | Mean | IL A: -27.97 : II A: -33.28: III B: -45.26: I	Grp | Mean | IL A: -26.21: III A: -35.25: II B: -49.90: I	Grp | Mean | IL A: -75.74 : III B: -95.42: II B: -96.88: I	Grp | Mean | IL A: -29.64 : III A: -35.93: II B: -50.04: I	Grp | Mean | IL A: 0.0139 : I A: -0.0417: III A: -0.9444: II

For all lifting tasks besides 20 lbs front lifting, at the destination position, the mean trunk flexion/extension torque of Integration Level II and III belongs to the same group, which is different from that of Integration Level I. For the 20 lbs front lifting, the mean values trunk flexion/extension torque of Integration Level I and II belongs to the same group, which is different from that of Integration Level III. For the trunk rotation torque value in side lifting task, there is no significant difference (F= 2.19, $p = 0.114$, df

=2) among the three Integration Levels. Figure 26 is the boxplot of the destination position trunk torques of all lifting tasks.

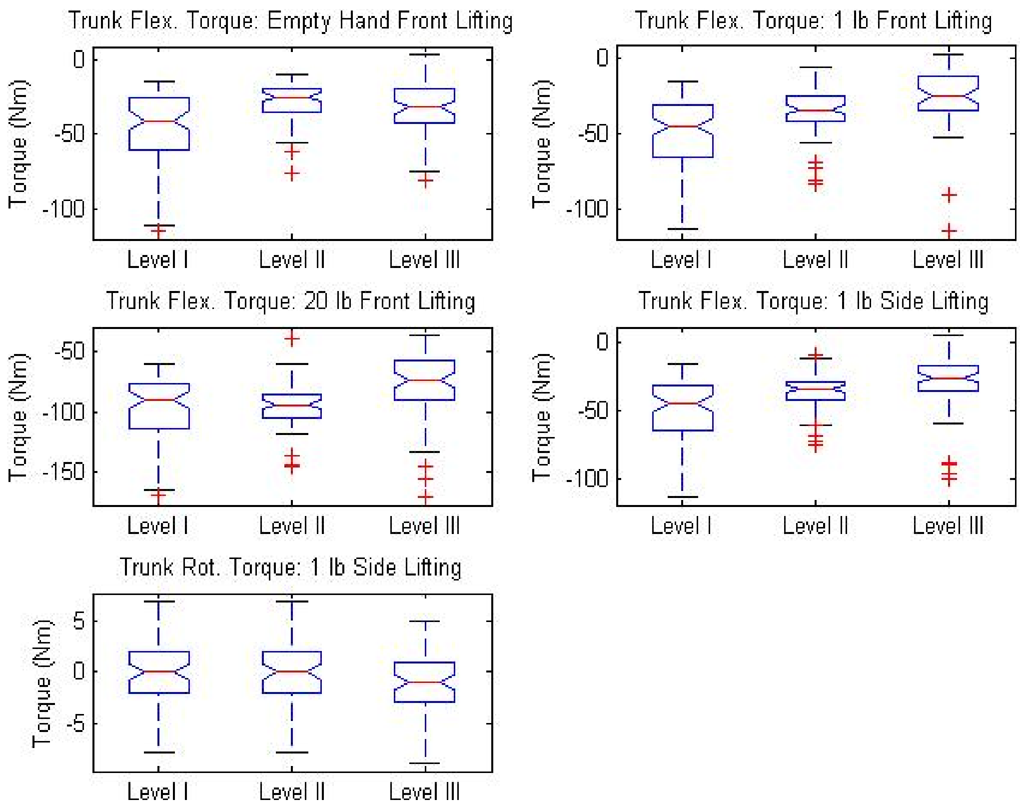

Figure 26 BOXPLOT OF THE TRUNK TORQUE: DESTINATION POSITION

5.4.4.2.2 Pushing Task

The Static Strength Prediction assessment is used to analyze the pushing task. Subject put his/her palm on the tool cart handler and push it forward. The horizontal pushing force is 2.25 kg. The shoulder torque and the elbow torque are chosen as the dependent variables. One-way ANOVA is performed to test whether there is significant difference of the mean value of the shoulder torque and elbow torque among the three Integration Levels. Following table is the one-way ANOVA results.

Table 24. ANOVA: SHOULDER, ELBOW TORQUE: PUSHING

	Pushing Task	
	Shoulder Torque	Elbow Torque
ANOVA	$F = 16.71$ $p < 0.001$ $df = 2$	$F = 0.21$ $p = 0.814$ $df = 2$
Fisher LSD	Group | Mean | IL A: 8.00 : II A: 7.73: I B: 5.46: III	Group | Mean | IL A : 1.4722: II A : 1.3403: I A : 1.2917: III

Figure 27 is the boxplot of the shoulder and elbow torques of pushing task.

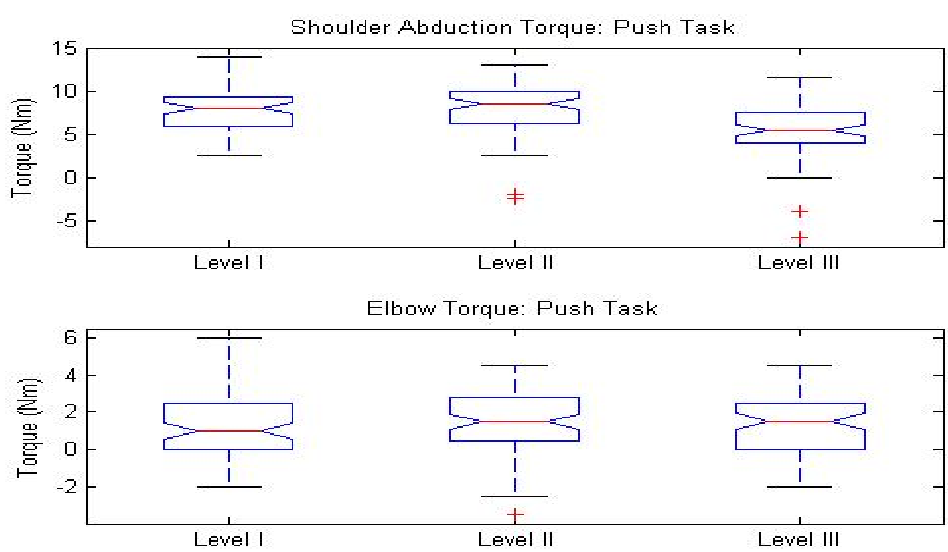

Figure 27 BOX PLOT OF SHOULDER, ELBOW TORQUE: PUSHING

The Integration Level has a significant effect on the shoulder torque ($F = 16.71$ $p < 0.001$ $df = 2$), but not elbow torque ($F = 0.21$ $p = 0.814$ $df = 2$). Fisher LSD test shows that the shoulder abduction torque of Integration Level I and II belong to the same group, which is different from that of Integration level III..

CHAPTER VI

CONCLUSIONS

This study focuses on validating the Virtual Build methodology for ergonomic assessments. The over-time reliability and criteria-related validity are tested. This study provides the practical justification of introducing Virtual Build into ergonomic research. The Virtual Build Integration Level I and II were demonstrated high validity and excellent reliability in conducting ergonomic assessments. The Integration Level III is reliable, but it needs some improvements to be valid.

6.1 Reliability

The correlation of the two ergonomic assessment score between two trials is used to evaluate the reliability performance of Virtual Build. For Integration Level I, all test-retest reliability scores are higher than 0.8. It tells that the Virtual Build in Integration Level I has "excellent" over-time reliability in conducting those ergonomic assessments. For Integration Level II, all test-retest reliability scores are higher than 0.8. It tells that the Virtual Build in Integration Level II has "excellent" over-time reliability in conducting those ergonomic assessments. For Integration Level III, 3 out of 17 test-retest reliability scores are lower than 0.8. It tells that the Virtual Build in Integration Level III has "Excellent" over-time reliability in conducting most ergonomic assessments, and "Good" over-time reliability performance in certain situation. Following table summarizes the reliability score in Virtual Build Integration Level I, II and III.

Table 25. SUMMARY OF RELIABILITY SCORE

Task	1991 NIOSH Lifting Equation	SSP					RULA
		Trunk Torque			Shoulder Torque	Elbow Torque	
	RWL		Flexion	Rotation			
0 lb Front Lifting	I: 0.8664 II: 0.8342 III:0.7589	Original Position	I:0.8988 II:0.8790 III:0.9504	N/A	N/A	N/A	N/A
		Destination Position	I:0.9667 II:0.9491 III:0.9048	N/A			
1 lb Front Lifting	I:0.9291 II:0.8793 III:0.7990	Original Position	I:0.9859 II:0.9207 III:0.9061	N/A	N/A	N/A	N/A
		Destination Position	I:0.9694 II:0.9534 III:0.8458	N/A			
20 lbs Front Lifting	I:0.9250 II:0.8159 III:0.8159	Original Position	I:0.9673 II:0.8243 III:0.9450	N/A	N/A	N/A	N/A
		Destination Position	I:0.9753 II:0.8579 III:0.8641	N/A			
Side Lifting	I:0.8567 II:0.8689 III:0.8230	Original Position	I:0.9015 II:0.9196 III:0.9664	I:0.8760 II:0.8358 II:0.7762	N/A	N/A	N/A
		Destination Position	I:0.9814 II:0.9517 III:0.9093	I:0.9626 II:0.9518 II:0.8601			
Pushing	N/A	N/A	N/A	N/A	I:0.9147 II:0.9000 II:0.9195	I:0.9701 II:0.8618 III:0.9314	N/A
Reaching	N/A	N/A	N/A	N/A	N/A	N/A	I: 1.0 II: 1.0 III: 1.0

Figure 28 plots the reliability score of all three Integration Levels.

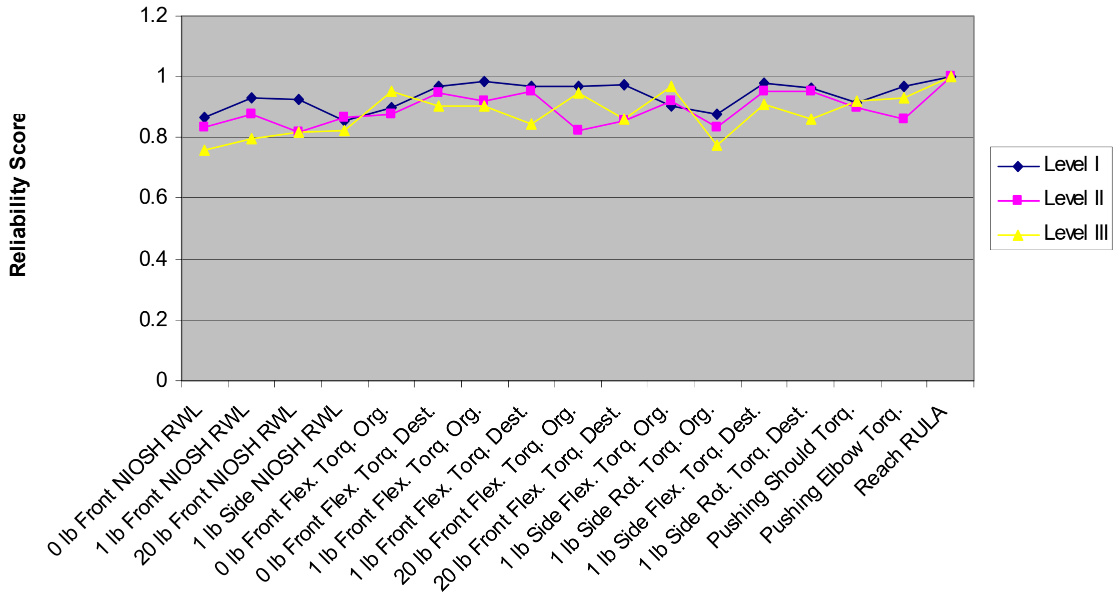

Figure 28 PLOT OF RELIABILITY INDEX

To test whether the Integration Level has a significant effect on the Virtual Build over-time reliability, a one-way ANOVA is performed with null hypothesis

H0: $\mu_I = \mu_{II} = \mu_{III}$ (no effect of Integration Level)

H1: *H0* is false

Table 26. ANOVA: RELIABILITY INDEX

ANOVA: Single Factor						
SUMMARY						
Groups	*Count*	*Sum*	*Average*	*Variance*		
Level I	17	15.94695	0.938056	0.002059		
Level II	17	15.20341	0.894318	0.002949		
Level III	17	14.97302	0.880766	0.004918		
ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	0.03048	2	0.01524	4.606063	0.014794	3.190727
Within Groups	0.158816	48	0.003309			
Total	0.189296	50				

The ANOVA shows that the Integration Level does have significant effect on the Virtual Build's over-time reliability performance ($F = 4.6061$, $p = 0.014794$, $df = 2$). It shows that the Integration Level I has higher over-time reliability performance than other two Integration Levels.

Generally, all three Integration Levels has mean reliability score higher than 0.8. This shows that, overall, all Integration Levels of Virtual Build can provide an "Excellent" over-time reliability in conducting ergonomic assessments. The Integration Level III has "Good" over-time reliability performance for some assessments at certain occasion. There are some works need to be done on Virtual Build Integration Level III to make it more over-time reliable for conducting ergonomic assessments.

6.2 Validity

Validity is evaluated by comparing the 1991 NIOSH Lifting Equation RWL score with the manual-measured and calculated result, which is assumed as error-free criteria result. The t test result shows that the Virtual Build Integration Level I is valid for the1991 NIOSH Lifting Equation assessment. The Virtual Build Integration Level II is valid for the 1991 NIOSH Lifting Equation assessment, when the human subject has a feedback from the box. The Virtual Build Integration Level III is not valid for 1991 NIOSH Lifting Equation assessment, with or without the feedback of box. The losses of perception and lack of feedback in the virtual environment may cause inaccurate motion of human subjects, while ergonomic assessments require accurate position, posture and motion information.

Following figure is the box plot of the 1991 NIOSH Lifting Equation RWL results of the four lifting tasks in all three Virtual Build Integration Levels. Each small figure represents one lifting task.

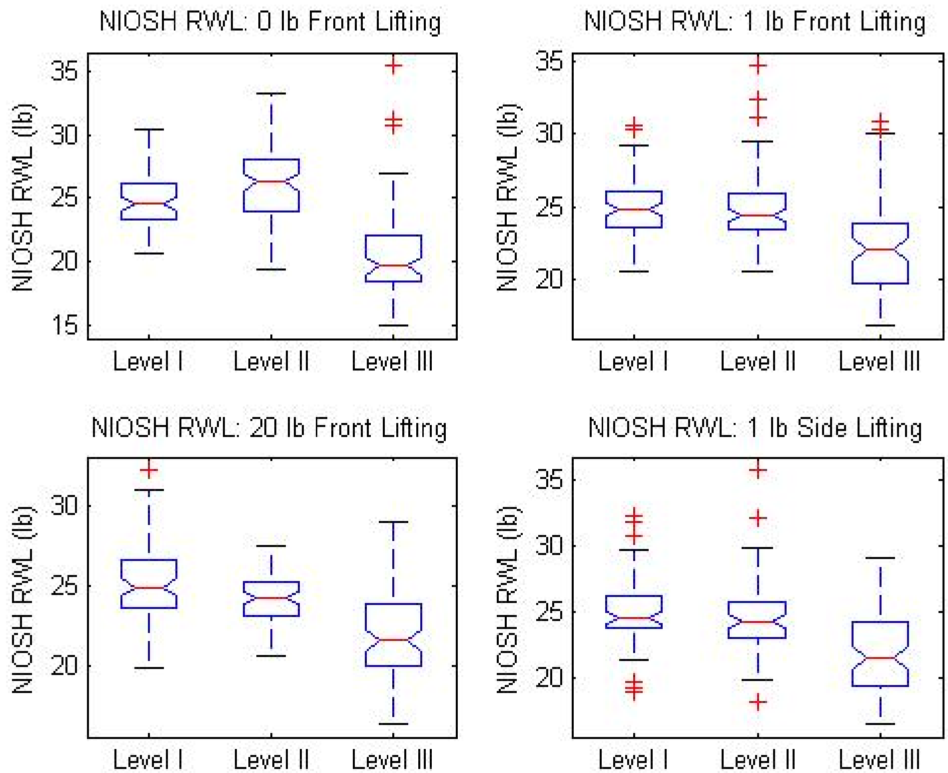

Figure 29 BOXPLOT OF NIOSH LIFTING EQUATION RWL

The comparison of ergonomic assessments result across Integration Levels shows that the Integration Level of Virtual Build has significant effect on ergonomic assessments. The interaction of Integration Level and External Loading has significant effect on 1991 NIOSH Lifting Equation assessments. The 1991 NIOSH Lifting Equation RWL score of Virtual Build using virtual environment is significantly different from that of Virtual Build using real environment or DHM only. For the Static Strength Prediction (SSP) assessment, the difference among three Integration Levels of Virtual Build is dependent on the task and body part that is analyzed. At the original position of lifting tasks, the trunk flexion torque of using Virtual Build Integration Level III is significantly different from that of either Level I or II, while there is no significant difference in the trunk rotation torque among three Integration Levels. At the destination position of lifting tasks, the trunk flexion torque of using Virtual Build Integration Level I is significantly different from that of either Level II or III, while there is no significant different in the trunk rotation torque among three Integration Levels. For pushing task, the shoulder torque of using Virtual Build Integration Level III is significantly different from that of either Level I or II, while there is no difference in the elbow torque among three Integration Levels.

CHAPTER VII

DISCUSSIONS

7.1 Real vs. Virtual

In this study, Virtual Build methodology is used as the tool to perform the ergonomic assessments. Mainly, two kinds of ergonomic assessments are chosen in this study. 1991 NIOSH Lifting Equation takes information about the position of certain part of human body as input, and Static Strength Prediction takes information of posture as input. The validity and reliability test of these two ergonomic assessments result by using Virtual Build can reveal whether the Virtual Build methodology is suitable for the position- or posture- determined ergonomic assessments.

For the 1991 NIOSH Lifting Equation assessment, the Integration Level I (DHM only) has the smallest variance and RMSE in 1991 NIOSH Lifting Equation RWL result among all three Integration Levels. As the researcher operates the digital manikin model to perform the task, the External Loading does not impact the final 1991 NIOSH Lifting Equation RWL result. The Integration Level III (DHM + MOCAP + VE) has bigger variance and error in 1991 NIOSH Lifting Equation RWL result than Integration Levels II (DHM + MOCAP + Mockup). It shows that human subject's perception of the location or position in virtual environment varies more than that in real environment. For both Virtual Build Integration II and Integration III, the additional perception feedback from the box in hand have a significant improvement on human subject's perception of location or position ($F = 12.20$, $p < 0.0001$, $df = 2$; $F = 5.32$, $p = 0.006$, $df = 2$). But the different weight level of box (1 lb or 20 lbs) does not have significant effect on the subject's judgment on location and position.

Static Strength Prediction assessment takes Posture and External Loading as inputs. This study does not cover the comparison of Static Strength Prediction result between different External Loadings. Static Strength Prediction assessment is used to analyze both lifting task and pushing task. For lifting task, postures at original position and destination position are assessed.

At the original position of all four lifting task, the mean of trunk flexion torque in Integration Level III is significantly bigger than that of Integration Level I and II ($F = 50.24$, $p < 0.001$, $df = 2$; $F = 28.93$, $p < 0.001$, $df = 2$; $F = 33.88$, $p < 0.001$, $df = 2$; $F = 66.21$, $p < 0.001$, $df = 2$).

Following pictures shows a subject in the original posture of the lifting task in the real environment.

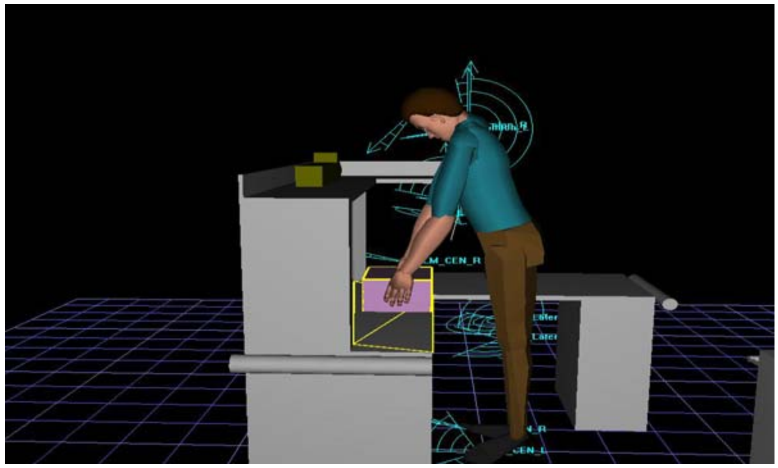

Figure 30 LIFTING IN REAL ENVIRONMENT: ORIGINAL POSITION

Following pictures shows a subject in the original posture of the lifting task in the virtual environment.

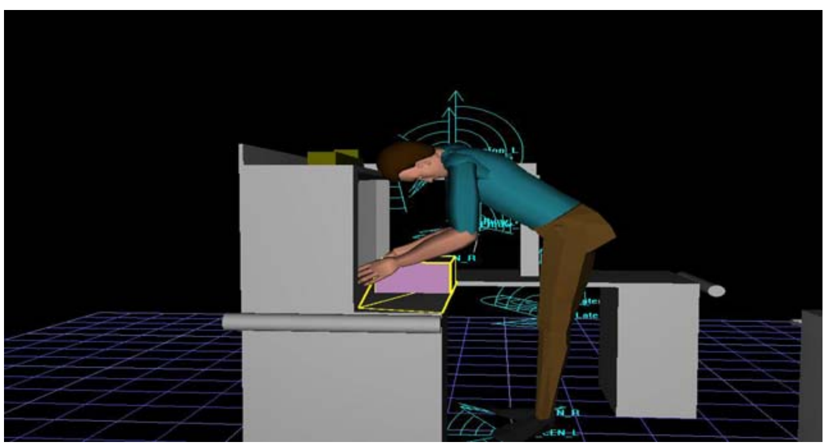

Figure 31 LIFTING IN VIRTUAL ENVIRONMENT: ORIGINAL POSITION

Most subjects bend more in virtual environment (Level III) than they do in real environment (Level II). The Figure 29 shows the posture of original lifting position in real environment, and Figure 30 shows the posture in virtual environment. The possible reason may be the field-of-view of the Head Mounted Display. The 5DT HMD only provides 40°diagonal field-of-view. Compared with the human normal field-of-view ranged from 160°to 208°, the Head Mounted Display field-of-view is quite limited. And at the original position of lifting, the box is out of the narrow view of Head Mounted Display, subjects adjusted their posture to locate the box.

At the destination position of lifting task, the box is in the normal view range and the head is return to the neutral position. There is no significant difference in the trunk torque between the virtual environment and real environment. Following pictures shows a subject in the destination posture of the lifting task in the real environment.

.

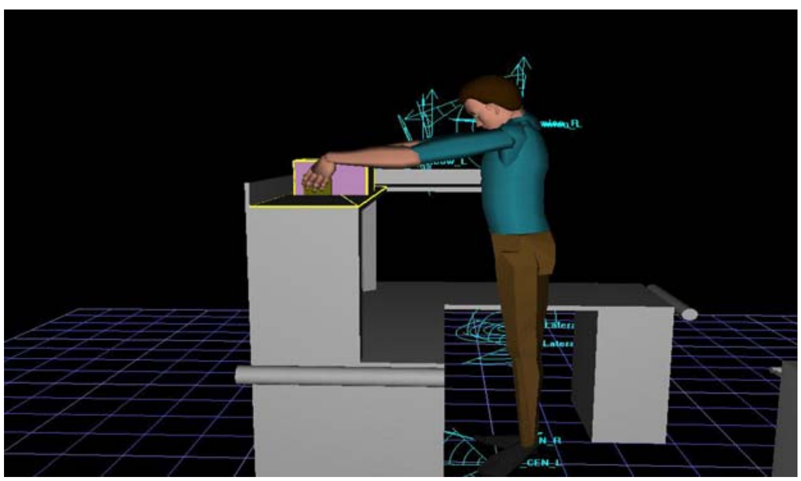

Figure 32 LIFTING IN REAL ENVIRONMENT: DESTINATION POSITION

Following pictures shows a subject in the destination posture of the lifting task in the virtual environment

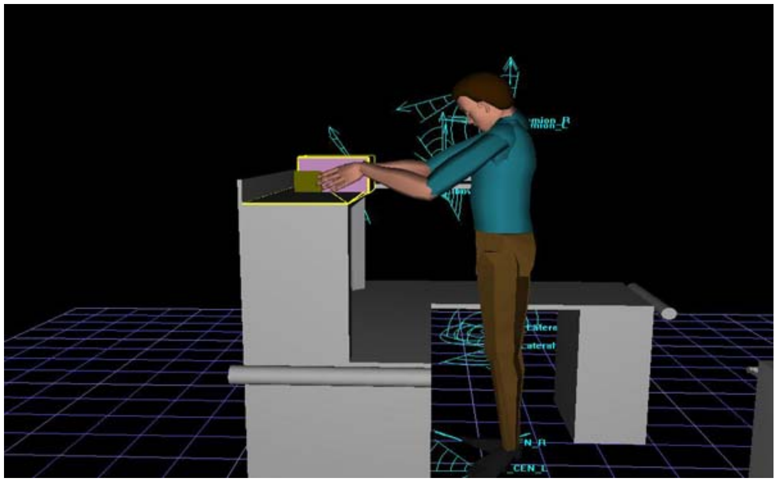

Figure 33 LIFTING IN VIRTUAL ENVIRONMENT: DESTINATION POSITION

7.2 External Loading

During the front lifting task, three different External Loading levels are set up. For the Integration Level I: which implementing DHM simulation only, the three External Loading levels does not make difference because the digital manikin model is manipulated by the researchers, who has no feeling or feedback from the external loadings. This is represented by the blue line with start marker in figure 20. For the Integration Level II, which uses the real environment, the External Loading levels are represented by the real weight of the box which subjects actually lift. 0 lbs lifting is lifting with hand emptied. 1 lb and 20 lbs lifting are lifting with box with 1 lb and 20 lbs weight respectively. For the Integration Level III, which uses the virtual environment, the External Loading levels are represented by the view that subjects can see and the instruction researchers give to subjects. During 0 lb lifting, subject can not see box in the virtual eye view. During the 1 lb and 20 lbs lifting, the subject can see a box in the virtual eye view. The researcher gives subjects instruction of imagining the box is either 1 lb or 20 lbs. From table 27, there is no significant difference of 1991 NIOSH Lifting Equation RWL result between External Loading of 1 lb and 20 lbs. There is significant difference of 1991 NIOSH Lifting Equation RWL result between External Loading 0 lb and 1 lb or 20 lbs($F = 12.20$, $p < 0.001$, $df = 2$; $F = 5.32$, $p = 0.006$, $df = 2$). From figure 20, the lines show for 0 lb lifting task, the 1991 NIOSH Lifting Equation RWL results of Integration Level II and III has wider span, and with the help of box (1 lb or 20 lbs), the NIOWH RWL result get closer, but there still significant difference between Integration Level II and III ($F = 34.15$, $p < 0.001$, $df = 8$). For all lifting tasks in Integration Level III, subjects actually do lifting with hand empty. There is no other feedback of the external loading except the eye view and instruction. The two feedbacks in virtual environment do not provide subjects enough cues, so that subjects can perform close to real environment. The result suggests that additional feedback is necessary to improve the Virtual Build Integration Level III.

7.3 System Reliability

Reliability is an internal characteristic of, instead of an input to, the Virtual Build system. The correlation between reliability and the system output error can reveal how the system's performance relies on the reliability. The system output error is evaluated by the Root Mean Square Error and the output deviation. Correlation tests between the reliability index (*ICC* score) and the Root Mean Square Error as well as the output deviation is conducted. Following table is the Pearson's r correlation score of ICC with the standard deviation and RMSE.

Table 27. CORRELATION BETWEEN RELIABILITY AND DEVIATION, RMSE

	Standard Deviation (*σ*)	RMSE
Reliability (*ICC*)	γ = -0.61763 *p = 0.0324*	γ = -0.73324 *p = 0.0067*

Figure 34 plots the standard deviation against the system reliability, also the linear fit line.

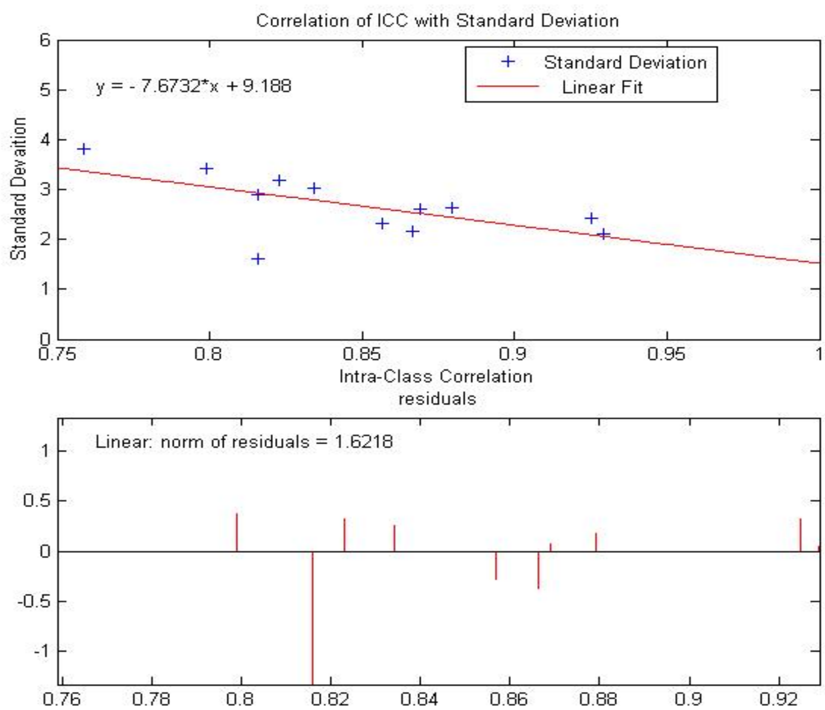

Figure 34 REGRESSION OF RELIABILITY AND STANDARD DEVIATION

The *F* statistic for the overall model is significant (*F*=5.40, *p*=0.0453), indicating that the system reliability (*ICC* score) explains a significant portion of the variation in the ergonomic assessments by using Virtual Build. The R-Square of 0.3748 indicates that system reliability (*ICC* score) accounts for 37.48% of the variation in Virtual Build ergonomic assessments. The downside of the regression represents the negative correlation between the system reliability with the Virtual Build output variation.

Figure 35 plots the RMSE against system reliability index and the linear fit line.

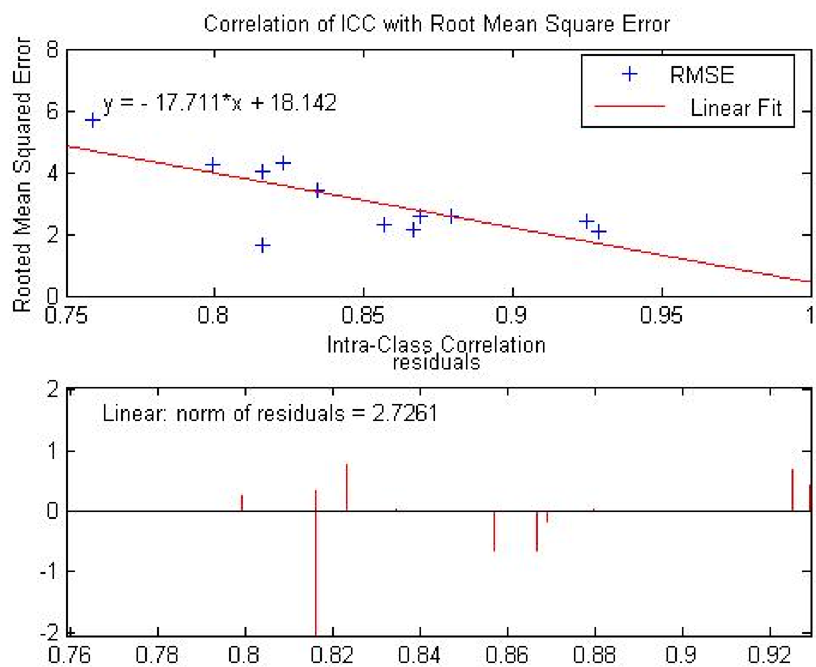

Figure 35 REGRESSION OF RELIABILITY AND RMSE

The *F* statistic for the overall model is highly significant (*F*=10.41, *p*=0.0104), indicating that the system reliability (*ICC* score) explains a significant portion of the error in the Virtual Build ergonomic assessments. The R-Square of 0.5363 indicates that system reliability (*ICC* score) accounts for 53.63% of the error in Virtual Build ergonomic assessments. The downside of the regression represents the negative correlation between the system reliability with the Virtual Build output error.

7.4 Limitation

One limitation of this study is that all tasks were conducted in one time session. The random errors from the motion capture suit, location of reflective markers and motion capture system calibration were not taken into consideration. If the two trials are taken in two different time session, the random factors, such as the motion capture calibration and markers' placement, can be included in. This may affect the result of this study.

Another possible limitation is the sample size. For reliability research, some researchers suggested sample size up to 400 participants (Charter, 1999). The 36 sample size in this study was partially justified by the previous 7 subjects experiment result. The increase of sample size may contribute to strength the result, especially for results in Integration Level III, which have bigger variances.

7.5 Future Work

The limited field-of-view of the Head Mounted Display may be the cause of the difference of Static Strength Prediction score between Integration II and III. In the virtual environment, with a narrower view, subject needs to adjust his/her posture differently from real environment to find the view. Increase the field-of-view of the Virtual

Environment equipment may make subject's behavior more realistic in virtual environment.

Edwards (2004) suggested that the force feedback should not be included in VE for ergonomic study if performance is evaluated by time. For the ergonomic study which is not evaluated by the performance time, this conclusion is not sustained. From Figure 20, we can find out that, for Integration Level III, with the help of a visible box, human subject perform lifting task more realistic than without visible box in hand. In the virtual environment, the 1991 NIOSH Lifting Equation RWL result get more close to real environment when the subject can have visual feedback. But this feedback itself is not enough to make Virtual Build Integration III valid. The inclusion of force feedback into Virtual Build Integration Level III is suggested in the future work.

All ergonomic assessments used in this study are static analyses, which are basically based on static posture. It a direction of future works to test the performance of using Virtual Build to conduct some dynamic ergonomic assessments.

REFERENCE

Aaras, A., Veierod, M., Laren, S., Ortengren, R., and Ro, O., 1996. Reproducibility and stability of normalized EMG measurement on musculus trapezius, *Ergonomics*, 39 (2), pp. 171 –185.

Badler, N., 1997. Virtual Humans for Animation, Ergonomics, and Simulation, *Nonrigid and Articulated Motion Workshop, 1997 Proceedings, IEEE.* 16 June.

Bartko, J., 1966. The intra-class correlation coefficient as a measure of reliability, *Psychol rep*. 19, pp. 3 – 11.

Bland, J. and Altman, D., 1986. Statistical methods of assessing agreement between tow methods of clinical measurement, *Lancet*, 8, pp. 307 – 310.

Brazier, J., 2003. The Car that Jill Built: A Case Study of the 2005 Mustang, *The 6th Applied Ergonomics Conference Proceeding*, Dallas, TX, March.

Brazier, J., 2004. An Ergonomic Success Story; The Launch of Ford Motor Company's 2004 F-150, *The 7th Applied Ergonomics Conference Proceeding*, Orlando, FL, March.

Bureau of Labor Statistics, 1997. *Workplace Injuries and Illness in 1995,* March.

Carey, M.B., and Koenig, R.H., 1991. Reliability assessment based on accelerated degradation: a case study, *IEEE trans, Reliability*, Vol. 40, pp. 499-506.

Carmines, E.G., and Aeller, R.A., 1979. Reliability and Validity Assessment, SAGE publication, London.

Cerney, M.M., Vance, J.M., and Duncan J.R., 2002. Using Population Data and Immersive Virtual Reality for Ergonomics Design of Operator Workstations, *SAE Digital Human Modeling Conference Processing*, pp. 107-120, June.

Cerney, M.M., Vance, J.M., and Duncan J.R., 2003. An immersive workstation design tool using three-dimensional anthropometric data, *Proceedings of the 47 Annual Meeting of the Human Factors and Ergonomics Society*, October, Denver, CO.

Chaffin, D.B., 1984. Occupational Biomechanics, Wiley, New York.

Chaffin, D.B., Faraway, J.J., Zhang, X.D., and Woolley, C., 2000. Stature, Age and Gender effects on reach motion postures, *Human Factors*, vol. 42, No. 3, pp. 408 – 420.

Chaffin, D.B., (Ed.) 2001. Digital Human Modeling for Vehicle and Workplace Design, Society of Automotive Engineers, Warrendale, PA: Society of Automotive Engineers.

Chaffin, D.B., 2002. On Simulating Human Reach Motions for Ergonomics Analyses, *Human Factors and Ergonomics in Manufacturing*, Vol. 12 (3), pp.235-247.

Chaffin, D.B., 2003 a. Digital Human Modeling and Simulation – Opportunities and Challenges, *International Ergonomics Association Conference Proceeding*, Seoul, Korea.

Chaffin, D.B., 2003 b. Improving Digital Human Modeling for Proactive Ergonomics in Design, *International Ergonomics Association Conference Proceeding*, Seoul, Korea.

Chaffin, D.B., 2005. Human Motion Simulation for Vehicle and Workspace Design, *the 8th Applied Ergonomics conference Proceeding*, New Orleans, LA, March.

Corlett, E.N., 1999. Rapid Upper Limb Assessment (RULA), from *The Occupational Ergonomics Handbook* (Ed. by Karwowski,W.; Marras, W.S.).

Dankaerts, W., O'Sullivan,P.B., Burnett, A.F., Straker, L.M., and Danneels, L.A., 2004. Reliability of EMG measurements for trunk muscles during maximal and sub-maximal voluntary isometric contractions in healthy controls and CLBP patients • *Journal of Electromyography and Kinesiology*, Volume 14, Issue 3, pp. 333-342 .

Davies, R.C., 1997. Affordable Virtual Reality in Ergonomics –some examples, *Proceeding of the 13 Triennial Congress of the International Ergonomics Association*, Tampere, Finland, 5, pp.89-91.

Delaney B., 1998. On the Trail of the shadow women: The Mystery of Motion Capture, *Computer Graphics and Applications, IEEE*, Vol. (18), 5, pp.14-19.

Denegar, C., and Ball, D., 1993 Assessing reliability and precision of measurement: An introduction to intra-class correlation and standard error of measurement, *Journal of Sport Rehabilitation*, 2, pp.35 – 42.

Edwards, G. W., Barfield, W., and Nassbaum, M.A., 2004.The use of force feedback and auditory cues for performance of an assembly tasks in an immersive virtual environment, *Virtual Reality,* 7: pp.112 –119.

Ehara, Y., Fujimoto, H., Miyazaki, S., Tanaka, S., and Yamamoto, S., 1995. Technical Note: Comparison of the performance of 3D camera systems, *Gait & Posture*, Vol. 3, pp. 166-169.

Ehara, Y., Fujimoto, H., Miyazaki, S., Mochimaru, M., Tanaka, S., and Yamamoto, S., 1997. Comparison of the performance of 3D camera system II, *Gait & Posture*, 5, pp.251-255.

Wertheim, E.H., Paxton, S. J. and Tilgner, L., 2004. Test–retest reliability and construct validity of Contour Drawing Rating Scale scores in a sample of early adolescent girls , *Body Image*, Volume 1, Issue 2, May , pp. 199-205 .

Fagarasanu, M., and Kumar, S., 2002. Measurement instrument and data collection: a consideration of constructs and biases in ergonomics research, *International Journal of Industrial Ergonomics*, 30(6), pp. 355-369.

Faraway, J.J., 2000. Modeling reach motions using functional regression analysis, *SAE Technical Paper* 2000-01-2175, Warrendale, PA.

Farsway, J.J., and Hu, J., 2001. Modeling variability in reach motions, *SAE Digital Human Modeling Conference*, Arlington, VA, 2001-01-2094.

Faraway, J.J., 2003. Data-based Motion Prediction, *SAE Digital Human Modeling Conference*, Montreal, Canada, 2003-0102229.

Fleiss, J., 1986. The Design and Analysis of Clinical Experiments, New York: John Wiley & Sons.

Flisher A.J., Evans, J., Muller, M. and Lombard, C., 2004. Brief Report: Test-retest of self-reported adolescent risk behaviour, *Journal of Adolescence*, vol. 27, Issue 2, April, pp. 207 – 212.

Freivalds, A., 2004. Biomechanics of the upper limbs, Taylor & Francis.

Frey, H.C., Mokhtari, A., and Zheng, J., 2004. Recommended practice regarding selection, application, and interpretation of sensitivity analysis methods applied to food safety process risk models. *Report for Office of Risk Assessment and Cost-Benefit Analysis,* U.S. Department of Agriculture, Washington, DC. Jan. 30.

Frisiello, S., Gazaille, A., O'Halloran, J., Palmer, ML., and Waugh, D., 1994. Test-retest reliability of eccentric peak torque values for shoulder medial and lateral rotation using the Biodex isokinetic dynamometer. *The Journal of Orthopaedic and Sports Physical Therapy*, 19(6), pp.341-344.

Gill.S.A., and Ruddle, R.A., 1998. Using Virtual Humans to Solve Real Ergonomic Design Problems, *International Conference on SIMULATION*, Sep, Conference Publication No. 457.

Gleicher M., and Ferrier N., 2002. Evaluating Video-based motion capture, *proceedings of computer animation*.

Hager, K. M.R., 2003. *Reliability of Fatigue Measures in an Overhead Work Task: A study of Shoulder Muscle Electromyography and Perceived Discomfort,* Unpublished M.S. thesis, Virginia Polytechnic Institute and State University.

Handbook for the economic analysis of health sector projects, 2000. Project Economic Evaluation Division, Economics and Development Resource Center.

Haslegrave, C.M., Baker, A., and Dillon, S., 1992. Use of Ergonomic workspace modeling in vehicle design. *Proceeding of International Conference on Computer-Aided Ergonomics and Safety* '92 – CASE '92, Tampere, Filand.

Henriksen, M., Lund, H., Moe-Nilssen, R., Bliddal, H. and Danneskiod-Samsoe, B., 2004. Test–retest reliability of trunk accelerometric gait analysis , *Gait & Posture*, Volume 19, Issue 3, pp. 288-297.

Kanis, H., 2000. Questioning validity in the area of ergonomics/human factors, *Ergonomics*, Vol. 43, No., 12, pp.1947-1965.

Keating, J., 1998. Unreliable inference from reliable measurement, *The Australian Journal of Physiotherapy*, 44, pp. 5 –10.

Kroemer K.H.E., 1999. Engineering Anthropometry, *The Occupational Ergonomics Handbook* (Ed. by Karwowski,W., and Marras,W.), CRC Press.

Landis, J., and Kock, G., 1977. The measurement of observer agreement for categorical data, *Biometrics*, 33, pp. 159 – 174.

Larrson, B., Karlssona, S., Erikssona M., and Gerdle, B., 2003. Test-retest reliability of EMG and peak torque during repetitive maximum concentric knee extensions, *Journal of Electromyography and Kinesiology*, 13(3), pp. 281- 287.

Laursen B., and Schibye B., 2002. The effect of different surfaces on biomechanical loading of shoulder and lumbar spine during pushing and pulling of two-wheeled containers, Applied Ergonomics, March, Vol. 33, No. 2, pp.167-174.

Li, K., Or, C., and Duffy, V.G., 2004 Universal Access and Workstation Design Using the Virtual Build Methodology, *The 7th Applied Ergonomics Conference Proceeding*, Orlando, FL, March.

Lok, B, Naik, S., Whitton, M., and Brooks, F.P., 2003. Effects of Handling Real Objects and Avator Fidelity on Cognitive Task Performance in Virtual Environment, *Proceedings of the IEEE Virtual Reality*.

McArdle, W.D., Frank, I.K., and Victor, L. K., 2001. *Exercise physiology: energy, nutrition, and human performance,* Philadelphia : Lippincott Williams & Wilkins.

Mital, A., Krueger, H., Kumar, S., Menozzi, M., and Fernandez, J.E., 1996. Redesign of lifting work using a 3-D Human Modeling software and the Revised NIOSH Lifting Equation", *Advances in Occupational Ergnonomics and Safety*, Amsterdam: IOS Press.

NIOSH, *Scientific Support Documentation for the Revised 1991 NIOSH Lifting Equation*: Technical Contract Reports, May 8.

NIOSH, 2001. *National Occupational Research Agenda for Musculoskeletal Disorders: Research Topics for the Next Decade*, A Report by the NORA Musculoskeletal Disorders Team, Jan.

Nunnally, J.C. 1978. *Psychometric Theory*, New York: McGraw-Hill.

Osman, A., Kopper, B.A., Barrios, F., Gutierrez, P. M. and Bagge, C.L. 2004. Reliability and Validity of the Beck Depression Inventory—II With Adolescent Psychiatric Inpatients, *Psychological Assessment*, Volume 16, Issue 2, pp. 120-132.

Panel on Musculoskeletal Disorders and the Workplace. Commission on Behavioral and Social Sciences and Education, 2001. *Musculoskeletal disorders and the workplace : low back and upper extremities*, Institute of Medicine (U.S.), National Research Council (U.S.) Washington, D.C. : National Academy Press.

Peters, G.A., and Peters, B.J., 2002. Automotive Vehicle Safety, Taylor & Francis, London.

Porter, J.M.; Case, K.; Freer, and Martin. T., 1999. Computer-Aided Design and Human Models, *The Occupational Ergonomics Handbook* (Ed. By Karwowski,W., and Marras, W.S.,) CRC Press.

Rider, K.A., Chaffin, D.B., Foulke, J.A., and Nebel, K.J., 2004. Analysis and Redesign of Battery Handling using Jack and HUMOSIM motions, *SAE Digital Human Modeling Conference Proceeding*, Michigan.

Robertson, G.G., Card, S.K., and Machinlay, J.D., 1993. Three views of virtual reality: Nonimmersive Virtual Reality, *Computer*, Vol. 26, issue 2, pp. 81-83.

Rosenblum, L. 2000. Virtual and Augmented reality 2020, *Computer Graphics and Application*. IEEE, Volume: 20, Issue; pp. 38-39.

Samuelsson, K.A.M., Tropp, H., Nylander, E., and Gerdle, B., 2004. The effect of rear-wheel position on seating ergonomics and mobility efficiency in wheelchair users with spinal cord injuries: A pilot study, *Journal of Rehabilitation Research & Development*, Vol. 41, No.1, pp. 65-74.

Saltelli, A. (Editor), Chan, K. (Editor), and Scott, E. M. (Editor), 2000. Sensitivity Analysis, John Wiley & Sons (October 15).

Saltelli,A., Tarantola,S., Campolongo, F., and Ratto, M., 2004. Sensitivity Analysis in Practice : A Guide to Assessing Scientific Models, John Wiley & Sons (April 9).

Shrout, P., and Fleiss, J., 1979. Intraclass Correlations: uses in Assessing Rater Reliability, *Psychological Bulletin*, 86(2), pp. 402-428.

Shrout, P., 1998. Measurement reliability and agreement in psychiatry. *Statistical methods in Medical Research*, 7, pp. 301-317.

Stanney, K.M (ed.) 2002. Handbook of Virtual Environments: design, implementation and application, Lawrence Erlbaum Associates, Mahwah, New Jersey.

Stanton, N., Young, M., 1997. Validation, The best-kept secret in Ergonomics, *Engineering psychology and cognitive ergonomics*, Brookfield, VT. 301-307.

Stokdijk, M., Beigstraatan, M., Ormel, W., Boer, Y., Veeger, H., and Rozing, P., 2000 Determining the optimal flexion-extension axis of the elbow in vivo – a study of interobserver and intraobserver reliability, *Journal of Biomechanics*, 33, pp.1130 – 1145.

Stone. R.J. 2002. Application of Virtual Environments: An overview, The *handbook of virtual environment* (Ed. By Stanney,K.M).

Tomović, R., 1963. Sensitivity analysis of dynamic systems, Translated by David Tornquist, New York, McGraw-Hill.

Tzannes, A., Paxinos, A., Callanan, M. and Murrell, G.A.C., 2004 An assessment of the interexaminer reliability of tests for shoulder instability, *Journal of Shoulder and Elbow Surgery*, Volume 13, Issue 1, January-February, pp. 18-23.

Waters, T.R., Putz-Anderson, V., and Garg, A., 1994. Applications Manual for the Revised NIOSH Lifting Equation, *Technical Information Service* (No. PB-94-110).

Waters, T.R., Putz-Anderson, V., Garg, A., and Fine, L.J., 1993. Revised NIOSH Equation for the Design and Evaluation of Manual Lifting Tasks, *Ergonomics*, 36(7), pp.749-776.

Wells, R., Moore, A.E., 1992. A Framework for Computer Assisted Approaches to the Prevention of Work-Related Musculoskeletal Disorders Involving Workplace Design and Modification, *Proceedings of the International Conference on Computer-Aided Ergonomics and Safety* '92 – CASE '92, Tampere, Filand.

Wilson, J.R., D'Cruz, M., Cobb, S. and Easgate, M., 1996. Virtual Reality for Industrial Applications, Nottingham University Press.

Youn, J.H., and Wohu, K., 1993. Realtime Collision Detection for Virtual Reality Applications, *Virtual Reality Annual International symposium,* 1993, IEEE, pp.18 – 22 Sept.

Yeung, S.S., Genaidy, A.M., Karwowski, W., and Leung, P.C., 2002. Reliability and Validity of self-reported assessment of exposure and outcome variable for manual lifting tasks: a preliminary investigation, *Applied Ergonomics* Vol. 33 pp. 463 – 469.

APPENDIX A

MOTION CAPTURE MARKERS LOCATION

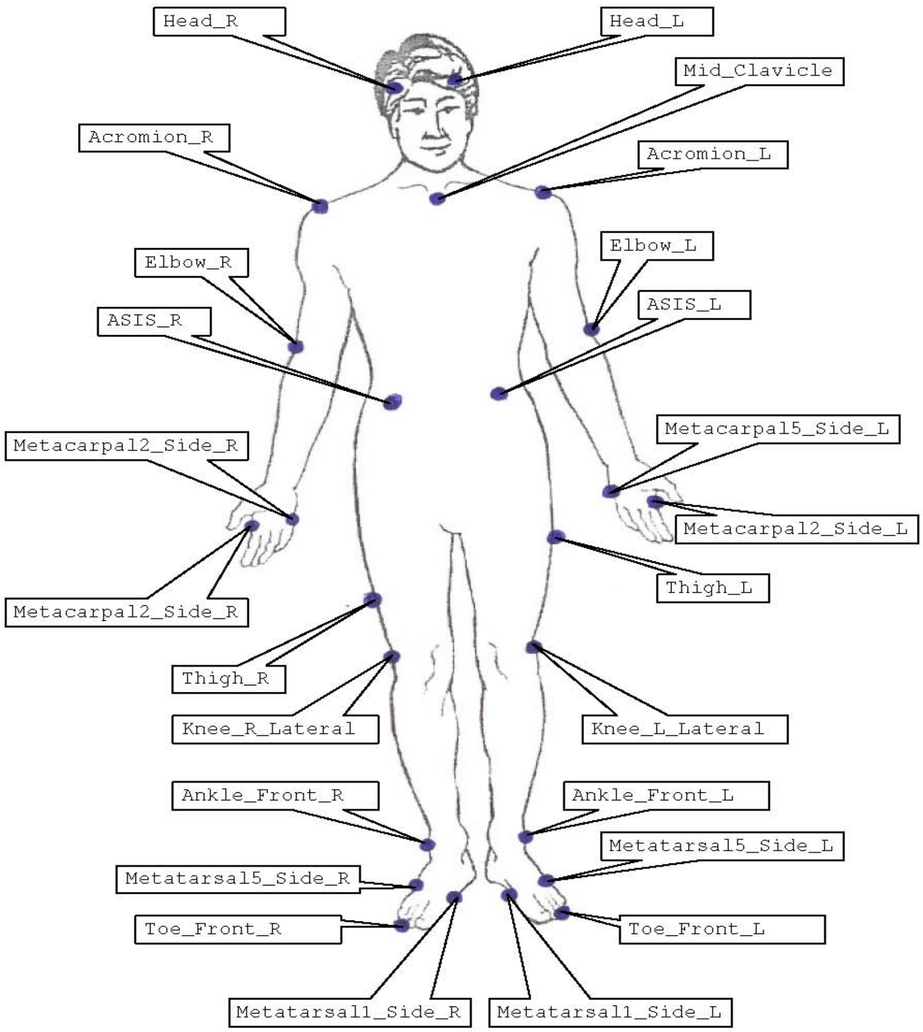

Figure 36 MARKERS FOR MOTIONANALYSIS SYSTEM (FRONT)

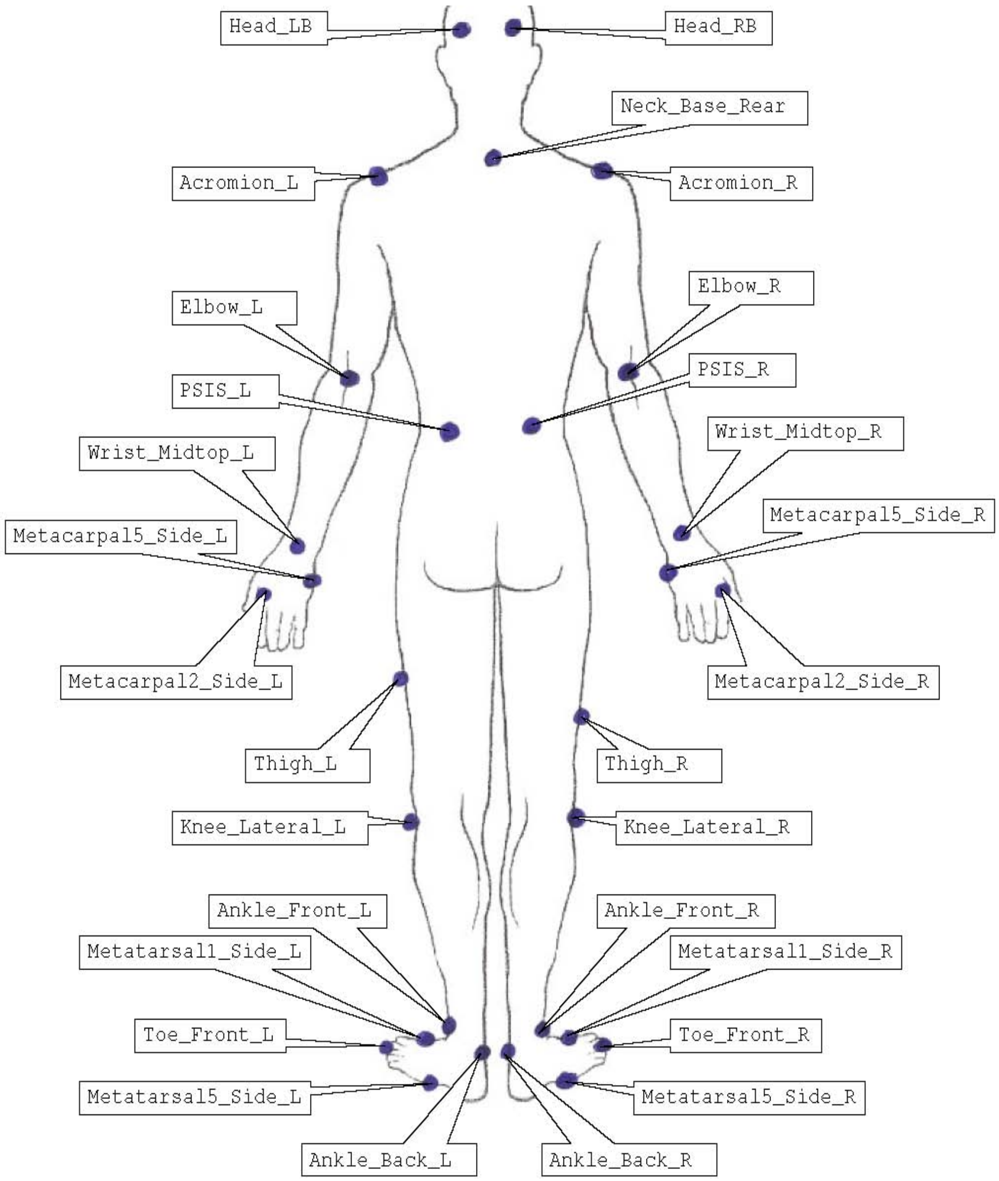

Figure 37 MARKERS FOR MOTIONANALYSIS SYSTEM (BACK)

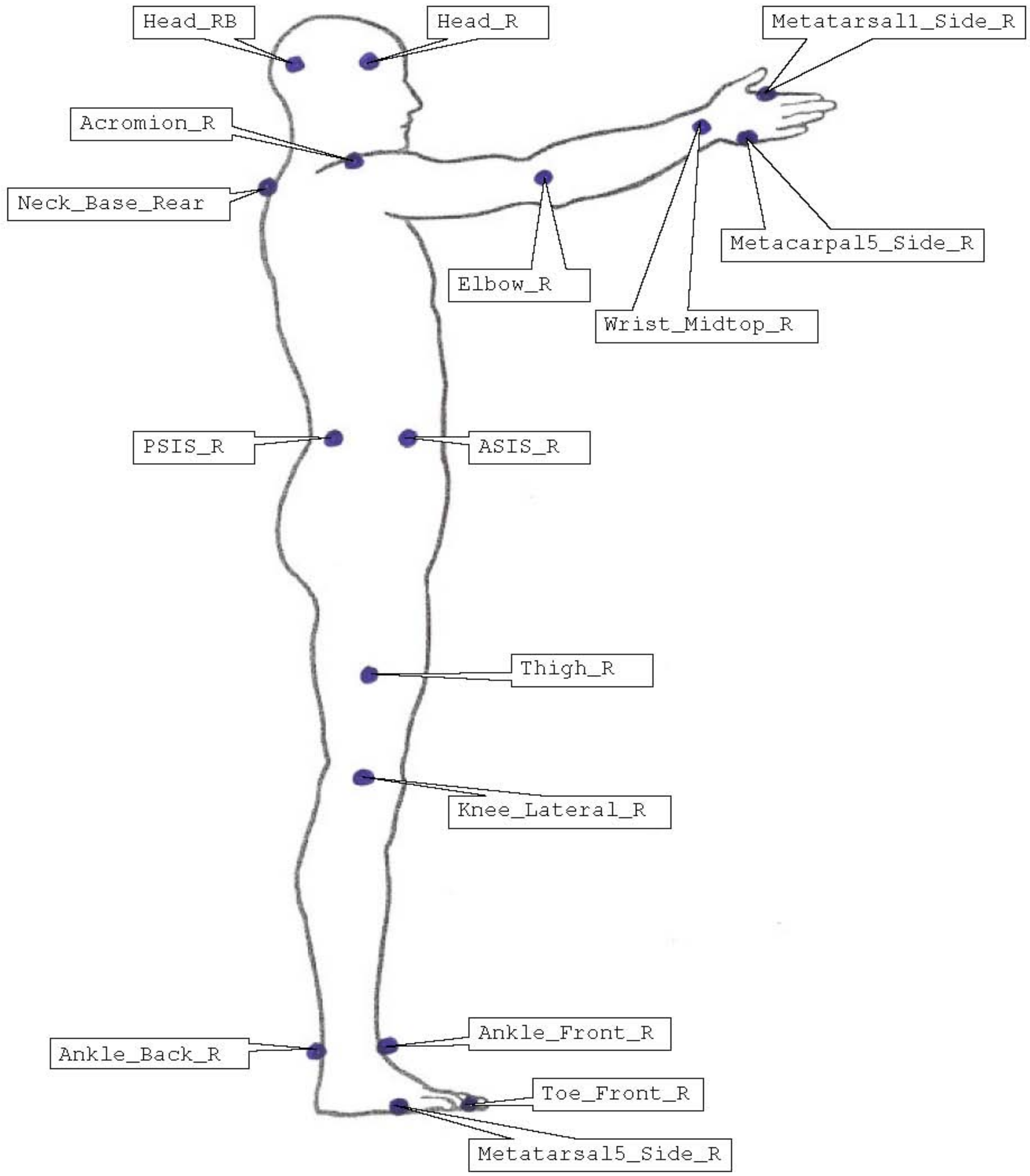

Figure 38 MARKERS FOR MOTIONANALYSIS SYSTEM (SIDE)

APPENDIX B

INFORM CONSENT FORM

INFORMED CONSENT for Participants in
"Validity and Reliability of Virtual Build for Ergonomics Assessments"

Principle Investigator: Dr. Vincent Duffy, Associate Professor, IE& CAVS at Mississippi State University

THE PURPOSE OF THIS STUDY: You are invited to participate in a research study to evaluate the usability of a new methodology for ergonomics called "Virtual Build". The experiment is designed to test the accuracy and consistency of the new methodology.

PROCEDURE:
To obtain this information, two experiments are to be conducted. The first experiment is conducted in real environment, and the second experiment is conducted in Virtual Environment. The procedures used in this study are as follows:

1. At first, we will take measurements of your body size, such as height, weight, arm length, and shoulder height.
2. You will wear a Motion Capture suit, which is used to track your motion. The procedure involves putting a special suit on your body which has reflective markers on key joint points (knees, elbows, etc) of the suit. The researchers will help you to put the suit on correctly.
3. The researchers will explain the data collection procedures to you. Also, the researchers will demonstrate the tasks that you will perform. These tasks include one front lifting task of 20 lbs, one front lifting with empty box (1 lb), one front lifting empty-handed, a side lifting with an empty box, a front reaching and a pushing. There are two trials of each task.
4. For the first experiment, you will perform tasks in a real environment, which includes a table, shelf, cart and box.
5. For the second experiment, you will wear a Head Mounted Display and perform the same tasks in a Virtual Environment.
6. After each experiment, we will ask a few questions about how difficult you think the tasks are.

The expected time of participation is 1 ½ hours.

RISK AND BENEFIT OF THIS RESEARCH: This study will help the researchers justify the next steps in the development of a new method for engineering design and ergonomics research. There is some small physical risk from the 20 lbs front lifting task and some risk of motion sickness from virtual environment exposure for this experiment. If you feel uncomfortable during testing, please notify the researchers immediately.

EXTENT OF ANONYMITY AND CONFIDENTIALITY: Individual identities will be protected and will not in any way be connected with any written summary of results that may later be published.

COMPENSATION: $25. The entire setup and experiment should take approximately 1 ½ hours.

FREEDOM TO WITHDRAW: You are free to withdraw from this study at any time for any reason without penalty.

APPROVAL OF THIS RESEARCH: The research project has been approved by the Institutional Review Board at Mississippi State University for projects involving human participants.

PARTICIPANT RESPONSIBILITIES: Participants should notify the researchers at any time if they want to stop participating in the study. Participants should also notify the researchers of any medical conditions that may interfere with results or increase of the risk of injury or illness.

PARTICIPANT'S PERMISSION: If you have any questions, please ask the researchers at this time.

I have read a description of this study and understand the nature of the research. I hereby consent to participate. I understand that I may discontinue participation without penalty at any time if I choose to do so.

Printed name: ___________________________

Signature & Date: ___________________________

For further information about this research, please contact: Dr. Vincent G. Duffy, CAVS & Department of Industrial Engineering, Mississippi State University, Mississippi State, MS 39762, (662) 325-5590, duffy@ie.msstate.edu

If you have additional question regarding your rights as a human participant in this research, you may contact the Mississippi State Regulatory Compliance Office at (662) 325-5220.