ABSTRACT

ZHANG, WENJUAN. Validation of a Virtual-Reality Based Psychomotor Test Using Physiological and Workload Measures (Under the direction of Dr. David B. Kaber.)

Psychomotor tests are human performance tests involving cognitive and motor processing. Some tests are widely used for cognitive and/or motor impairment diagnosis or for motor skill screening for occupational tasks. Typically, traditional psychomotor tests require well-trained professionals for task delivery, participant performance observation and performance rating. However, some critical issues have been identified with this process. First, the consistency of test delivery (or lack thereof) can be influenced by differences in communication skills of test administrators. Second, measurements of participant performance can also be influenced by administrator individual differences and personal opinions. Third, psychomotor tests place high attentional demands and cognitive workload for test administrators. Fourth, data collection is time consuming.

With advances in information technology, computerization of psychomotor tests has occurred to address some of the above limitations of physical test delivery. Some studies have been conducted in an attempt to validate the use of computerized versions of psychomotor tests and confirm their accuracy and reliability relative to native tasks with a focus on performance measures. Unfortunately, few, if any, studies have used biometric or workload measures as bases for validation of such systems. The present study sought to validate a previously developed virtual reality (VR)-based simulation of the block design (BD) task incorporating a haptic control interface. Task condition (native vs. VR) was used as a subject-grouping variable. Twenty-four participants were recruited and half were assigned to the VR condition. These participants received haptic device training, including 40

trials of a virtual dice-manipulation task. Participants under the native task condition directly began testing. In order to develop a more complete understanding of any difference in mechanisms by which participants perform computerized vs. traditional versions of the BD task, electromyography (EMG; muscle activation level) and performance responses were collected during the test trials. The degree of mental workload imposed by the computerized version of the BD task, as compared to the native version, was also assessed by using the NASA-Task Load indeX (TLX). Finally, correlation analyses were performed on the various response measures in order to identify any similar trends.

Results revealed the intensity of muscle use to be comparable across the native and VR versions of the BD task. However, task completion time with the VR simulation was still higher than that with native BD task. Results on the mental workload assessment also showed comparable levels of mental demand and physical demand among the native and VR test conditions as well as comparable performance ratings. However, ratings of temporal demand, frustration and overall effort, as well as the overall NASA-TLX score, were significantly different among the test conditions. Finally, significant correlations were observed among response measures, including the EMG responses with task time and workload ratings. Different patterns of response correlation were found under different phases of movement (honing vs. rotation) in the BD task as well as the different test conditions (native vs. VR).

These findings indicate that it is possible to develop computerized-versions of standardized psychomotor tests that cause users to exhibit similar types and levels of motor behavior as in native task performance, and that computerized tests lead to similar levels of mental and physical demand. On this basis, VR simulations of psychomotor tests may be used to facilitate objective and systematic evaluation of motor control disabilities as well as to provide a platform for training and learning of fine motor skills. Such computer-based simulations may help users achieve similar levels of skill rehabilitation as native tasks. Similarly, the simulation investigated in the present study may support motor skill development in abled persons for complex training of assembly tasks in industrial environments.

Validation of a Virtual-Reality Based Psychomotor Test Using Physiological and Workload Measures

by
Wenjuan Zhang

A thesis submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the Degree of
Master of Science

Industrial Engineering

Raleigh, North Carolina

2015

APPROVED BY:

David B. Kaber, PhD
Committee Chair

Manida Swangnetr, PhD

Peter Bloomfield, PhD

BIOGRAPHY

Wenjuan Zhang enrolled at Beijing University of Aeronautics and Astronautics (BUAA) in 2009 and studied engineering science in the program of École Centrale de Pékin. She graduated in 2013 with a Bachelor of Science in Mathematics and Applied Mathematics. In the same year, Wenjuan was admitted in the Master of Industrial Engineering program at North Carolina State University. In her second year, she developed interest in Human Factors and Ergonomics from her research experience and changed her program to Master of Science in Industrial Engineering. She was also employed as a teaching assistant for related courses at the undergraduate and graduate levels. Upon completion of her degree, she plans on continuing his studies as part of the Doctorate program in Industrial and Systems Engineering at NC State.

ACKNOWLEDGMENTS

I deeply appreciate the help and support of my advisor Dr. David Kaber, who always generously provided guidance and extremely constructive feedback. I would like to thank Dr. Manida Swangnetr and Dr. Peter Bloomfield for serving on my committee. Dr. Swangnetr's familiarity with EMG and Dr. Bloomfield's expertise in statistical analysis methods helped to form a rigorous and meaningful thesis. I'd also like to acknowledge my colleagues in the ergonomics lab, especially Wenqi Ma, who was always willing to listen to me and share her experience. Without any of these people, this research would not have been possible.

TABLE OF CONTENTS

LIST OF TABLES vii

LIST OF FIGURES viii

LIST OF ABBREVIATIONS ix

1. Introduction 1

1.1 Native Psychomotor Tests 1

1.2 Computerization of Psychomotor Tests 6

2. Literature Review 10

2.1 Validation of Computerized Psychomotor Tests 10

2.2 Virtual Reality Version of Block Design (BD) Test 16

2.3 EMG as Tool for Analysis of Motor Task Performance 24

2.4 Cognitive Workload Assessment of Computerized Psychomotor Tests 32

2.4.1 Primary and secondary task performance measure 34

2.4.2 Physiological measures 35

2.4.3 Subjective measures 39

3. Problem Statement 43

3.1 Summary of literature review 43

3.2 Research Expectations 44

4. Method 45

4.1 Tasks.......... 45

4.2 Apparatus 46

4.2.1 Native BD task.......... 46

4.2.2 Computerized BD task.......... 47

4.2.3 EMG system 49

4.3 Participants.......... 50

4.4 Independent variables.......... 50

4.5 Dependent variables 51

4.6 Experiment design.......... 51

4.7 Procedures 52

4.8 Research hypotheses 54

5. Data analysis 55

5.1 Diagnostics and transformation.......... 55

5.2 EMG responses 55

5.3 Task completion time.......... 56

5.4 Workload (NASA-TLX) 57

5.5 Correlation among response measures.......... 57

6. Results.......... 57

6.1. Muscle activation 57

6.2 Task completion time 61

6.3 Subjective measure of workload 62

6.4 Correlation analyses on response measures 64

7. Discussion **..** 70

8. Conclusion 80

8.1 Applications 81

8.2 Limitations 82

8.3 Future work 83

REFERENCES.. 85

APPENDICES.. 92

Appendix A Informed Consent Form.. 93

Appendix B Experiment Procedure... 96

LIST OF TABLES

Table 6.1 Descriptive statistics summary for EMG responses (% of MVC) 58

Table 6.2 MANOVA results to test trial and phase effect 58

Table 6.3 MANOVA results on test condition effect 59

Table 6.4 Two-way ANOVA results on test condition and muscle effects within phase of movement 60

Table 6.5 Mean (SD) summary for BD time-to-task completion time (sec.) 61

Table 6.6 Results of ANCOVA test for time-to-task completion 62

Table 6.7 Mean (SD) summary of TLX overall score and ratings of sub-dimensions 63

Table 6.8 Results of two-sample t-tests on TLX score and sub-dimension ratings 64

Table 6.9 Correlation of time-to-task completion and EMG 65

Table 6.10 Correlation of NASA-TLX scores and EMG under native condition 67

Table 6.11 Correlation of NASA-TLX scores and EMG under VR condition 68

Table 6.12 Correlation of NASA-TLX scores and time-to-task completion 69

LIST OF FIGURES

Figure 4.1 Custom workstation for native BD test 47

Figure 4.2 Virtual reality workstation for Block Design test.. 48

Figure 4.3 VR-based BD test simulation 49

Figure 6.1 Results of post-hoc analysis on EMG responses.. 61

Figure 6.2 Task completion time of native and VR conditions across trials 62

LIST OF ABBREVIATIONS

3D	Three-dimensional
AMA	Averaged Muscle Activations
ANOVA	Analysis of Variance
ANCOVA	Analysis of Covariance
BD	Block Design
(BD)	Blink Duration
BR	Blink Rate
BQSS	Boston Qualitative Scoring System
CH	Channel
CFF	Critical Flicker Fusion
DOF	Degrees of Freedom
DRAT	Discriminant Reaction Time Tasks
ECG	Electrocardiographic
EEG	Electroencephalogram
EMG	Electromyography
EOG	Electrooculographic
ERP	Event-related Potential
ESS	Equivalent Set Size
GEFT	Group Embedded Figures Test
HRV	Heart-rate Variability

HUD	Head-up Display
IVIS	In-Vehicle Information Systems
MCH	Modified Cooper-Harper
MVC	Maximum Voluntary Contraction
MWL	Mental Workload
NASA-TLX	NASA Task Load Index
OI	Orientation Index
PC	Perceptual Cohesiveness
PPT	Purdue Pegboard Task
PSD	Power Spectrum Density
RE	Real Environment
RPE	Rated Perceived Exertion
ROCF	Rey-Osterreith Complex Figure
RULA	Rapid Upper Limb Assessment
SWAT	Subjective workload assessment technique
TBI	Traumatic Brain Injury
TSS	Total Set Size
TTC	Time-to-task Completion
TTT	Total Task Time
VE	Virtual Environment
VR	Virtual Reality
WASI	Wechsler Abbreviated Scale for Intelligence

Introduction

1.1 Native Psychomotor Tests

Psychomotor tests, or human performance tests involving cognitive and motor processing, have been in existence for some time (Causby et al., 2014). Such tests are used to diagnose cognitive and/or motor impairments, serve as mechanisms for skill training, and predict psychomotor abilities (Stefanidis et al., 2006). Some tests have been widely used in clinical studies (Causby et al., 2014) as well as in industry in order to select qualified workers for certain jobs (Hunter and Hunter, 1984).

There are several classical psychomotor tests that have been used most frequently by by researchers. The Block Design (BD) test is a psychomotor task in which participants are required to arrange colored blocks in order to construct an identified pattern. It requires the use of spatial ability in positioning an orienting blocks in order to ensure that the combination of colored block faces accurately reproduces the target pattern. This task is most frequently presented as a subtest on the Wechsler Adult Intelligence Scale (WAIS) to measure visual-spatial abilities (Grote, 1986). Block Design reconstruction tasks have also been frequently used to diagnose organic brain damage (Goldstein & Scheerer, 1941; Kohs, 1923; Wechsler 1958; Schorr et al., 1982). It was found that the BD test distinguished between brain-damaged and normal individuals with high accuracy (80%). Spreen and Benton (1965) observed the BD test to be one of the most effective measures among behavioral tests designed for this purpose (Spreen & Benton, 1965). (It has also been found that the BD test reliably predicts outcome in a brain-injury rehabilitation program (Diller et al., 1974).) It was suggested that brain-damaged individuals do not perform well on the BD test because of an inability to adopt an analytic approach (logical and analysis-based) in pattern reconstruction. In the BD test, the identified design pattern must be mentally segmented into units corresponding to colored block faces, then the blocks are directly placed to match each unit of a pattern. Alternatively, participants may pursue a “synthetic” approach (trial and error) in which blocks are manipulated until the work product matches the entire pattern (Schorr et al., 1982). Schorr et al., (1982) examined able participant strategy in the BD task through various aspects of performance. Three experiments found a predominant analytic strategy. The time to place a single test block in a workspace decreased with the greater the number of interior edges against which to match the block in the design pattern. For example, block faces with cross-sectional coloring, requiring an orientation judgment, were placed slower than solid block faces. A fourth experiment predicted overall construction times for a design from the number of solid blocks and interior edges available for matching blocks. Schorr et al. recommended that such strategy analysis is necessary to assess behavior and performance in other neuropsychological tests. Individual differences have also been found to play a role in BD test performance. Older adults were substantially slower and less efficient compared to younger adults (Salthouse, 1987). The level of complexity in the BD task was found to be manipulated in terms of three variables: the total block set size (TSS); equivalent set size (ESS); and perceptual cohesiveness (PC). ESS was defined as the number of equivalent alternatives obtained when a design is turned through successive 90 rotations and PC was calculated by counting the number of adjacent same-color edges. Therefore, it is important to account for these influential task factors as well as individual differences in terms of applications and research with the BD task.

In addition to the BD task, other psychomotor tasks have been developed for neuropsychological tests, including the Rey-Osterreith Complex Figure (ROCF). The ROCF is widely used for evaluation of visuospatial constructional ability and visual memory (Shin et al., 2006). It was developed by Rey in 1941 and standardized by Osterrieth in 1944. The ROCF task involves copying a complex geometric figure and then reproducing it from memory, either immediately, following a delay, or both. The test provides data on individual neuropsychological functions, including attention and concentration levels, fine-motor coordination, visuospatial perception, nonverbal memory and organizational skill (insert refs.), as well as brain dysfunctions (insert refs.). Analysis of ROCF data provides the capability to diagnose and evaluate participant motor disability. For example, Gallagher and Burke (2007) conducted a study with the ROCF involving copy and recall trials. The copy trial was used to assess the perceptual analytic process or strategy an individual used to complete the figure. The recall trial drawing was scored for accuracy as an index of visual memory. Gallagher & Burke (2007) demonstrated that age, gender and IQ all had significant effects on raw ROCF scores for memory trials. Luzzi et al. (2011) obtained similar results, including effects of age and education on copying, immediate recall and delayed recall as well as on the rate of figure forgetting. Other empirical research has demonstrated that the ROCF reflects executive functions, such as organizational and planning abilities, in addition to visual cognitive functions due to the complexity of the figure (Ogino et al., 2008).

There are a variety of methods for scoring ROCF copies or reproducitons. Many methods are based on Taylor's (year) 36-point system and assess the presence and accuracy of each drawn element of the ROCF. The Boston Qualitative Scoring System (BQSS) was developed to evaluate qualitative features of the ROCF as well as unit presence and accuracy. Ogino et al. (2008) applied the BQSS measure on ROCF reproductions by 78 children with various neuropsychological disorders and found significant correlation with results of other tests, including scores on maze completion. Scoring of the ROCF has also been demonstrated to be a method for rapid assessment of participant health states. For example, Rocha-Amador et al. (2009) used the ROCF test to evaluate visuospatial organization (copy trials) and visual memory (immediate recall trials) in children living in areas of Mexico exposed to different mixtures of neurotoxic agents. The ROCF was found to be effective for rapid quantitative measurement of health risks.

Another psychomotor task, used to assess motor ability and cognitive processing, is the Purdue Pegboard test. This test was originally developed to provide a basis for selection of employees for industrial jobs requiring manipulative dexterity (Tiffin & Asher, 1948). It was designed to incorporate the desirable features of numerous dexterity tests into a simple and easily administered performance test. The PPT provides separate measurements of the right hand, left hand, and both hands together. It measures two different abilities; dexterity for gross movements of hand, fingers and arms as well as fine hand motor dexterity needed in small assembly work. The PPT has been identified as the most widely used tests for fine finger dexterity (Sartorio et al., 2013). The PPT requires participants to place pins in holes within a time limit using their left hand, right hand or both hands respectively. The PPT also involves an assembly subtask. Participants are instructed to use their right hand to pick-up one pin from a cup and place it in a top hole on the right side of the test board while (at the same time) using the left hand to pick-up a washer. As soon as the pin has been placed, the participant is instructed to drop the washer over the pin.

The PPT has the advantages of easy and rapid administration. As a result, it is an effective and widely used research instrument. Corney (1953) used the assembly subtest as part of the PPT to investigate how well manual dexterity scores predicted manual dexterity in other tasks. Corney examined performance of a pair of individuals and found the PPT to be predictive of dexterity.

The PPT has also been used with patients for identification of different injury states or disabilities. Costa et al. (1963) found the PPT to be a useful, short-sensory motor test, independent of brain-injury patient education level. They also found the test to be effective in detecting and screening for the presence of brain damage (laterality lesions). This result was supported by Kane and Richard (1972). They attempted to determine the value of the PPT for screening children for learning disabilities through public schools and other agencies that deal with large numbers of kids but have time and personnel constraints for such activities. They also suggested that the Pegboard might have value as a simple and objective measure for identifying benefits of classroom instruction for children with learning disabilities.

The PPT has also been used in some recent research to investigate fine motor control in right- and left-handers in the absence of practice in order to compare and contrast uni-manual and bi-manual performance (Judge and Stirling, 2003). This study found that left-handers perform more proficiently than right-handers on the assembly component of the PPT, which requires co-ordination of both hands. They also stated that the assembly task of the PPT differentiated quite well between right- and left- handers. Even more recently, Telles et al. (2012) used the PPT as a technique to assess manual dexterity and eye-hand co-ordination under varied breathing conditions. They suggested that the immediate effect of alternate nostril breathing included improvements in PPT performance requiring attention, bimanual dexterity and visuo-motor co-ordination.

1.2 Computerization of Psychomotor Tests

Traditionally, all of the above psychomotor tests are performed physically, with at least one administrator and appropriate test instruments. General requirements include well trained clinical staff and/or lab researchers, who are able to proficiently deliver the tests. Several specific issues have been identified with physical psychomotor test performance. First, consistent delivery of the test can be an issue. Although the methods to perform these tests are well defined, the way that each person administers the test may vary slightly. Factors such as clarity of explanation of the rules and communication skills of testing personnel are not always controllable and can cause influence participant performance. Second, measurements as part of tests can be influenced by administrator individual differences and personal opinions. Many tests involve monitoring task time and scoring output, which can vary from one evaluator to another. Use of multiple evaluators can lead to biased and inaccurate test results. For example, in the BD task, a test administrator must determine whether a participant has put a block in its precise place as part of a design or not. For the same block placement, it is possible that one administrator would consider the placement to be good enough to be called “successful” while another person would not agree. Third, the administration of such tests demands high attention of administrators. Related to this, the need to observe and monitor performance accurately can cause mental stress and fatigue.

Fourth, data collection is a tedious job. When administrating physical psychomotor tests with traditional apparatus, administers need to make notes on participant performance characteristics. Subsequently, this information is used as a basis for calculating task scores and assigning an overall evaluation based the data collected. Although software with calculation functions can facilitate this process to some degree, the work of data collection and analysis remains time consuming and laborious for most tests.

With advances in information technology, computerization of psychomotor tests has occurred to address some of the limitations of physical test delivery. In general, the use of technology has been pursued to make tests easier and more efficient to administer. Beyond this, computerization supports consistent delivery of tests and automated and accurate recording of response measures. For example, with computer presentation, it is possible to measure response latencies in the PPT task with millisecond accuracy, which is necessary for reaction time studies. Automatic data collection can reduce human labor in use of tests for analysis of able performance as well as disabilities.

Computerized tests and assessments have been developed in various research areas. Sahakian and Owen (1992) illustrated how developments in computing technology could be exploited for neuropsychological assessment of neurodegenerative diseases. They stated that computerized tests were sensitive to, and could distinguish between, neurodegenerative diseases even early in the course of a disease process. Lin and Wu (2013) explored the validity of computerized scaling of bilateral motor coordination in children 4–6 years of age. They found computer-based testing to be a valuable tool for identifying problems of bilateral motor coordination and for promoting early intervention to remedy these problems. Hertzog

(2014) conducted a computerized assessment of age differences in memory belief by using traditional paper-and-pencil versions of questionnaires vs. computer-based versions involving a mouse and visual analog scale ratings. They said participants found the scaling method easy to use and automated scoring was conducted on ratings. Berchou and Block (1983) used computerized psychomotor tests to identify central nervous system effects of drugs. They observed advantages of computer-controlled behavioral testing systems to include a high degree of test control, uniform test condition delivery, ease of test administration, and efficient data collection and analysis. Berchou and Block (1983) also observed that once such systems are setup, they require minimum skill to operate, and reduce the need for subjective evaluation and judgments by clinicians. Related to these findings, Towles (2004) observed that use of a virtual environment for psychomotor test delivery provides administrators with complete control over mechanical properties of the testing environment.

The above studies exploited the benefits of information technology for psychomotor test delivery and have provided some evidence of advantages of computerization. However, all prior validations of computer-based tests for research or clinical use have been limited to participant performance assessments. There is a need for validation of computerized tests by comparison of human physiological responses with responses during physical test performance. Such analysis might also provide insight into mechanisms behind participant performance and why similarities or differences exist among test types. In addition, while the use of computer technology may facilitate testing processes, it is possible that computer interfaces also pose additional performance challenges for participants. It is important to ensure that computerized tests do not create higher cognitive workload for participants than otherwise experienced in physical forms of tests, potentially causing fatigue. An assessment of user cognitive workload in performing physical and computerized psychomotor tasks would be helpful to address this issue. In summary, combined analysis of performance, physiological responses and cognitive workload for various physical and computerized formats of psychomotor tests would provide more substantial evidence of the validity of computerized tests.

2. Literature Review

2.1 Validation of Computerized Psychomotor Tests

Given the potential benefits of information technology, several studies have sought to validate computerization of standardized psychomotor tests. Berchou and Block (1983) described two interactive, computer-controlled tests, including critical flicker fusion (CFF) and discriminant reaction time tasks (DRAT). The purpose of their study was to examine whether the two tasks were sensitive to the behavioral effects of drugs, including diazepam, lorazepam and dextro-amphetamine, and to delineate the time course of drug effects. In the critical flicker fusion task, the stimulus presentation was computer controlled. On each of a series of trials, one of the diodes was flickering, while the other diode was constantly illuminated. Subjects were instructed to press one of two response switches, one on the left and one on the right, depending on whether the left or right diode was flickering. The discriminant reaction time task was also computer controlled. Subjects viewed a small screen on which a series of single digits was presented. Subjects were instructed to press a response button as quickly as possible each time the digit 4 was displayed. According to their result, both of the computerized tasks, the CFF and DRAT, proved sensitive to the impairment produced by 15 mg diazepam and proved sensitive in a dose-related fashion to the impairment produced by lorazepam. The authors stated that there was a continuous interaction between subject performance and the computer control of the administration of the task. This research identified a set of objective and quantifiable tests that are sensitive to drug effects, when administered in a highly controlled fashion. The tests are also brief in duration. The authors indicated that use of computerized assessment systems might prove a useful tool for a psychopharmacology research, especially with the increasing availability of inexpensive microprocessor systems.

Fillmore (2003) considered reliable and accurate assessment of human motor skill to be important to research fields and applied settings that concern diagnoses and rehabilitation of motor abnormalities. In his early research (Fillmore & Vogel-Sprott, 1994, 1995,1998), a computerized version of the pursuit rotor task was developed. Instead of holding a stylus to track the rotating target, participants tracked the target by moving a computer mouse on the desktop and controlling a visual targeting sight on a monitor. The objective of the task was to move the mouse to keep a circular, cross-hair locator sight in contact with the rotating target. Like the traditional tabletop pursuit rotor task, the computerized task measured performance as participant time-on-target during a test trial. The task was shown to be a reliable measure of motor skill, and comparisons with the traditional pursuit rotor task indicated convergent validity. In addition, the computerized task offered greater automation of operation and more programmable features. However, the computerized pursuit rotor task was developed on computer systems that were not compatible with Microsoft DOS or Microsoft Windows operating systems.

Fillmore (2003) subsequently introduced a new computerized version of the task that was developed for use under the Microsoft Windows operating systems. He performed two experiments with this version. The first one investigated the reliability of computerized assessment of psychomotor performance and the second investigated sensitivity of performance to alcohol-induced impairment. The first experiment measured acquisition and retention of participant psychomotor skill when they repeatedly performed the pursuit rotor task during two test sessions. The task stimuli were presented on a computer screen. As in the earlier computer version of the task, participants tracked the target by moving a mouse on a tabletop to control an on-screen circular, targeting cross-hair sight. The program automatically computed and stored test trial scores. The presentation of trials and rest intervals was automated. The program controlled the entire test procedure and allowed participants to perform the task alone in a study room. The experimenter was only with a participant in order to provide initial instruction and explanation of the procedure at the outset of the sessions. This experiment showed that the pursuit rotor task provided a reliable and stable measure of psychomotor performance. Analyses of performance showed participants gradually improved in performance during training trials, much like that observed using the earlier computerized version of the task. This improvement was retained after a 2-week period, and the test-retest reliability estimate showed that individual differences in achievement remained fairly consistent over that time.

Fillmore's second experiment tested the sensitivity of the pursuit rotor task as a measure of alcohol-induced behavioral impairment. The sensitivity of the task to functional impairment as a result of changes in blood alcohol concentration was assessed by testing performance at various times. After baseline trials, half the participants were randomly assigned to an alcohol treatment condition, while the other half received a placebo beverage. Participants then performed five blocks of trials as their blood alcohol concentration increased, peaked, and gradually declined. This experiment showed that participant psychomotor impairment was generally related to their blood alcohol concentration at each test block. The results of Fillmore's (2003) study provided additional support for the computerized pursuit rotor task as a reliable and stable measure of psychomotor performance that is sensitive to the disruptive effects of alcohol.

Virtual reality (VR) has also been used as a technology for computerization of psychomotor tasks. One major advantage of VR is that key aspects of real-world tasks can be simulated with a higher level of realism than simple 2-D computer-based simulation in order to more effectively support skill assessment and training through repeated trials (Kaber et al., 2014). The use of VR for this purpose can also reduce real training task costs as well as trainer workload. Li et al. (2010) developed a haptic-based virtual environment system for diagnosis and rehabilitation of Traumatic Brain Injury (TBI) patients. The VR system presented a virtual form of the Rey-Osterrieth complex figure (ROCF) reproduction test. This test involves participant copying or recreating from memory a standardized figure including 18 units of various shapes and segment types. Each unit is scored for accuracy or reproduction and placement. A hybrid-design interface was developed to provide participants with access to the task simulation in a manner similar to natural drawing behavior. The system hardware included a custom workstation with an embedded horizontal display area and integrated simulation projection system. The setup also included a desktop SensAble Technology Phantom haptic device that was placed on top of the workstation. The ROCF was projected on to the display area and participants were instructed to copy the figure in a work area. This haptic-based VR system was used for recording and scoring patient ROCF drawing tests. Patients completed the ROCF with the haptic device, and their performance was evaluated in terms of physical movements, time to completion, and drawing accuracy and placement (drawing score). Drawing performance was assessed based on a fuzzy model of handwriting. The system integrated haptic force rending, and patients were expected to be able to draw the ROCF with the haptic device as if they were using pen and paper. The system also integrated automated drawing unit pattern recognition and scoring. These capabilities improved the ROCF test efficiency and robustness. The author suggested that the computer-based system would significantly reduce the workload of neuropsychologists and that it could also be used as an open source tool for research in TBI patient testing and rehabilitation.

As an extension of Li et al. (2010) research, Kaber et al. (2014) evaluated use of the haptic-based VR simulation of ROCF reproduction as a basis for assessing motor skill training. They tested and trained 24 healthy participants in the standardized psychomotor control task using both native and VR formats. Participants were asked to use their non-dominant hands in order to promote sensitivity of the study for revealing any VR design features that might serve to accelerate motor learning. The experiment design replicated a simplified occupational therapy regimen and the ROCF test was identified to represent an occupational task, which was anticipated to improve as a result of therapy. Another VR simulation of the block design (BD) test (as part of the Wechsler Abbreviated Scale for Intelligence (WASI: The Psychological Corporation, 1999)) was developed to serve as the training task, which was expected to impact ROCF test performance. To perform the ROCF, participants used a SenSable Technologies Omni haptic device to virtually draw the complex figure elements on a horizontally aligned monitor while the simulation automatically recorded participant performance data. Results revealed significant improvements in test performance following training in the VR simulation of the BD test, including augmented haptic features. Training in the native form of the BD task, as well as a basic version of the VR simulation absent of enhanced haptic features, did not produce improvements in ROCF test performance. It is, however, important to note that performance during training was consistently better with the native task, but the enhanced VR simulation had greater impact on drawing task learning. This finding was attributed, in part, to a correspondence between the VR training and test task interface. In general, the findings of this study support use of VR-based haptic simulations of standardized psychomotor tests for motor skill training.

Pontonnier et al. (2014) assessed the capability of a VR-based assembly task simulation to evaluate physical risk factors in ergonomics. They designed and conducted experiments on simulated assembly tasks in real environments (RE) and virtual environments (VE). The VR system used was a high-resolution stereoscopic immersion room including a front-screen and a floor-screen. User interaction with the VR occurred through a Flystick2 (ART). The motion of the upper body and muscle activity from five muscles (comprising a kinematic chain) were recorded to in order to determine several objective indicators of risk exposure, including the Rapid Upper-Limb Assessment (RULA) score, averaged muscle activations (AMA), and total task time (TTT). Rated perceived exertion (RPE), using the Borg scale, and a questionnaire were also used as subjective indicators of risk exposure. The assembly task included several elementary operations and conditions similar to actual industrial processes: target reaching, object manipulation, piece sorting, standing posture, and repetitive motion. The authors manipulated task complexity (2 types of fitters vs. 6 types of fitters), timing regime (assemble as fast as possible vs. waiting for a sound signal before taking a new piece) and task environment (RE vs. VR) as within-subject factors. Results revealed significant differences between the measured indicators among the RE and VE. While objective measures indicated lower activity and exposure in the VE, the subjects experienced more discomfort than in the RE. A correlation of RULA scores across environments indicated that the trend of RULA scores for the VEs would be similar to that observed for the REs. Total task time was also found to be significantly higher in the VE than the RE, confirming a lack of familiarity of subjects with the VE. This study demonstrated the use of VR for ergonomic assessments of assembly task performance to be challenging. However, the high level of correlation between RE and VE results indicated that the VR-based simulator could be effectively used for such assessments.

2.2 Virtual Reality Version of Block Design (BD) Test

To date, several computer-based versions of the BD test have been developed. In a version developed by Salthouse (1987), a model pattern, a single block, and a 3x3 numbered grid were presented. Participants could use the arrow and "Enter" keys on a keyboard to rotate the block on the screen and place it instantly by pressing a number key (1-9) corresponding to a cell within the grid. Rozencwajg and Corroyer (2001) developed a software package called SAMUEL to simulate the BD test. In their simulation, the test design was displayed on the left of the screen and the reconstruction area was displayed on the right. A collection of six squares representing each of the six possible block placements was distributed across the bottom of the screen. Participants could use the mouse to "drag" the squares to the corresponding portion of the construction. The motivation for both these simulations was to collect additional human performance data (e.g., block manipulation patterns), which would be difficult to capture through direct observation of participant native task performance

(Salthouse, 1987; Rozencwajg & Corroyer, 2001; Clamann et al., 2013). However, as Clamann et al. (2013) stated, these versions of the BD test incurred a cost in terms of fidelity. In the first design, block movements were eliminated and participants simply conveyed blocks to the target grid by pressing buttons. In the other version, participants used a mouse to drag the blocks to desired locations, resulting in hand translation, but block rotations were completely eliminated. Moreover, in both of these computer-based versions, block placements were restricted to 90 degree rotation intervals (Clamann et al., 2013). This situation prevented participants from making rotation errors but also resulted in elimination of the last two design models in the WAIS-III and WASI BD protocols, which are rotated 45 degrees from the other designs.

Considering these limitations of existing computerized BD tests, previous research at North Carolina State University attempted to develop a higher fidelity version of the BD test that could record participant block movements and rotations in use of a single control. Clamann et al. (2013) developed a haptic-based VR-based simulation of the BD test, based on the WASI BD protocol. The VR simulation was presented on a PC with a stereo monitor using a NVIDIA® 3D VisionTM Kit, including active light-shutter goggles and an emitter. The VR features included a virtual tabletop divided into two parts: a work area and a display area. The display area presented the design to be replicated by the participant. The work area was used for block assembly. The haptic control interface was a SensAble Technologies PHANTOM Omni® Haptic Device, including a boom-mounted stylus providing 6 degrees-of-freedom (DOF) for movement and 3 DOF in force feedback. The Omni was used to manipulate a cursor appearing on the display. Participants could touch the cursor to a block and press and hold a button on the stylus to “grab” the block. The blocks could be lifted from the table’s surface and rotated along any axis using the Omni’s stylus. Once a block was returned to any area of the work surface, it was automatically released from the user’s “grasp”. Blocks on the work surface were presented as solid objects. Participants could feel a solid resistance when a block touched the work surface and a similar resistance when touching the cursor to a block or when two blocks collided.

Clamann et al. (2013) conducted an experiment to evaluate the VR-BD test and to compare performance with the native form of the task, including an evaluation of pre-test (baseline) performance, multiple training sessions, and a post-test to measure improvement. All participants received a baseline performance assessment with the WAIS-III BD test, using standardized materials. Participants were then required to complete eight WASI BD training trials distributed across three separate sessions (3 hours of total training time), as a physical task or using the VR, based on assigned condition. Participants were retested during the last session by using the WAIS-III BD test to identify any changes from baseline performance. Their results revealed that training in either the VR or the native versions of the task produced significant performance increases from baseline. However, neither version resulted in significant improvements over the other, suggesting that the VR-BD was at least as effective as the Native BD for supporting task training and follow-on performance. This VR-BD simulation was demonstrated to be effective for testing psychomotor skill, as compared to Native BD task performance.

Clamann et al. (2013) subsequently incorporated haptic, visual and a combination of haptic and visual aids in their VR simulation design in attempts to accelerate motor skill training through the BD test. They conducted an experiment to compare learning effects of the various aiding conditions, following motor skill training. Haptic aiding was designed as a snap force that pulled blocks to a target position during correct placement and presented rejection forces (acting against the block) during incorrect placement. Visual aiding provided feedback during incorrect block placements. If a user attempted to place a block in an incorrect orientation in the target grid, a yellow "X" or an arrow would be superimposed on the block. The "X" would appear when the wrong block face was showing. The arrow would appear when the correct block face was showing, but the block was rotated incorrectly. The arrow indicated the direction in which the block needed to be rotated for correct orientation in the target grid square. If a user moved a block to the correct orientation over the grid, those squares in the grid at which the block could be placed without error were highlighted in yellow. In addition to the passive visual assistance, subjects could request additional assistance during training. If the subject touched the control cursor to the design pattern at the top of the screen or the target grid, visual cues would be displayed on how to correctly place blocks. These cues indicated the orientation and locations of individual block faces. Touching the cursor to a target grid square would highlight the corresponding square in the stimulus pattern and any blocks in the workspace that matched the selected square. Likewise, touching the stimulus pattern would highlight the corresponding square on the target grid. The gridlines disappeared when any surface outside of the stimulus pattern or grid was contacted with the cursor. Participants performed a baseline evaluation prior to multiple training sessions and a post-training evaluation in order to identify any performance improvements. There were two test tasks anticipated to reveal improvements as a result of the “therapy” sessions, including the ROCF and BD subtest from the Wechsler Adult Intelligence Scale – Third Edition (WAIS-III; [14]). Both of these tests were administered once prior to training and once following the training sessions. The VR simulation with aids was expected to promote psychomotor performance over the basic simulation and the combination haptic and visual aiding was expected to lead to the greatest increase in motor performance. Results showed that the three conditions resulted in similar increases; however, on average the haptic condition showed the highest percentage of WAIS BD test improvement.

Clamann and Kaber (2012) also conducted another study to observe how haptic and visual assistance in the VR simulation of the BD task affected individual performance strategy selection. They used the haptic-based VR-BD that was also investigated by Clamann et al. (2013). Prior research broadly categorized strategies in the BD task as “analytic” or “global” (Rozencwajg & Corroyer, 2001; Schorr et al., 1982). Participants who employ an analytic strategy mentally segment each design into squares corresponding to block faces and place each block directly to match the squares. Participants employing a global strategy tend to view the design as a whole (gestalt or complete form), as opposed to individual squares. Clamann & Kaber (2012) expected field independent participants to be more likely to employ an analytic strategy and field dependent participants to be more like to employ a global strategy. Three types of aids, including visual, haptic and a combination of both (similar to those studied by Clamann et al., 2013), were implemented. It was also expected that the visual and combination forms of aiding would help participants implement analytic strategies. Prior to training, participants’ field dependence was assessed using the Group

Embedded Figures Test (GEFT). Subsequently, each participant visited the research lab for three separate training sessions, completing eight BD trials in total (10 designs per trial as required by the established WASI protocol). The primary dependent measure in the study was an orientation index (OI) defined as a measure of the number of tries required to place a block correctly as part of a pattern. A high OI implied an analytic strategy and low OI implied a global strategy, as an analytic strategy is associated with direct and structured design assembly while a global strategy is associated with trial and error. Analysis of the average OI across participants and trials for each condition indicated that visual aiding helped participants achieve an analytic strategy more so than the other conditions. However, regression analysis applied to the OI scores across training trials indicated that participants in the visual and combination conditions took advantage of the visual assistance early in training and maintained a consistent strategy (relying on the assistance) throughout the experiment. Participants in the haptic condition, in contrast, developed strategies independent of the aiding condition. Furthermore, the researchers stated that the OI may be a more sensitive measure of visuomotor learning in the BD task than the native task performance measure (the BD score). Haptic feedback alone was found to support participant analytical strategy development across several training trials comparable to the performance of participants making some use of visual assistance. It was concluded that haptic aiding appeared to support analytical strategy development over time; whereas visual aiding supported users in achieving such a strategy early in therapy and throughout performance. In addition to these finding, this study also confirmed the reliability of the VR-BD simulation for research use.

Using the same VR-BD task simulation, Jeon et al. (2013) investigated three-dimensional (3D) goal-directed movements. As a well-known approach to analysis of movement, Fitts' law can be used to accurately predict movement between a starting point and a target. However, this model provides little insight into the factors that account for variability during an overall movement. Jeon et al. analyzed sub-movement structure in order to further understand target acquisition movements in the BD task. Their study was also motivated by the capability of the VR system to capture high resolution kinematic data on design pattern reconstruction, which could provide learning insights beyond existing BD score interpretation. Jeon et al. used a stereoscopic display along with a haptic device with stylus control to present to participants the VR-BD simulation, based on the WASI. In their study, detailed kinematic data on the movements between grasping and releasing a block at the target grid and grasping another block were collected for analysis. These movements were assumed to be goal-directed since they required rapid aiming. Pauses were defined as movements in which the speed of the cursor remained below 5% of the overall peak movement speed and the movements occurring in between pauses were defined as sub-movements. Each participant visited the research lab for three consecutive days in which they completed eight BD test trials, including 10 designs per trial. The study results revealed that some of the kinematic measures, including overall sub-movement count (the number of sub-movements in an overall movement) and correction sub-movements (the number of sub-movements in a correction phase), were significantly correlated with BD scores. The kinematic measures on the correction phase were more strongly related to the BD scores than those on the ballistic phase (i.e., the number of sub-movements in the ballistic phase), suggesting that BD task performance was more dependent on fine motor skill. The authors interpreted the lack of perfect correlations among the kinematic measures with the psychomotor task scores as each kinematic measure accounting for specific aspects of performance (with measures on the ballistic phase describing gross motor skills and measures on the correction phases describing fine motor skills). They also found that a predictive model in kinematic parameters for the correction phase had a higher coefficient of determination (R^2), indicating the possibility to measure visuospatial and motor skills by analyzing kinematic measures on partial movements in a correction phase instead of using an elaborate standardized psychomotor test. In addition to these noteworthy results, their study also further supported the validity of the haptic-based VR simulation of the BD task for psychomotor task performance assessment.

In order to provide a more realistic experience with the VR-BD simulation, Clamann et al. (2014) made further modifications. Instead of using a stylus control interface, as in the previous research, a block shaped control was integrated with the Omni, replicating the experience of holding a block or cube. This modification was expected to cause training performance levels with the VR-BD task to more closely approximate those achieved in native BD task performance. Three groups of participants were recruited, including a Block group (using the block shaped control), a Stylus group (using the stylus control) and Control group (performing the BD native task). A LEGO assembly task was used to resemble an industrial assembly task for baseline performance testing and post-training tests. Participants in the Block or Stylus group completed the baseline performance evaluation, multiple training sessions in the native BD task, VR training using the block or stylus control, and post-training performance assessment. Participants in the Control group only completed baseline and post-training tests. Results indicated significant improvements in assembly test performance based on the VR-BD task training. However, improvements in task completion time were only significant for the Block and Control groups but not the Stylus condition. Moreover, the Block group produced greater improvements in assembly accuracy than both the Stylus and Control groups. While training with the block control had similar effects to the control group in terms of task completion time, the lack of significant improvement following stylus training suggests training with an incompatible device (relative to use of objects in the native task) may lead to negative skill transfer. It was concluded that VR-based training incorporating a control causing hand postures and motions similar to those used in the target task appeared to be effective for training assembly tasks.

2.3 EMG as Tool for Analysis of Motor Task Performance

In this section, literature is reviewed regarding how muscle activation analysis can be applied for assessing motor task performance. Electromyography (EMG) is a well-accepted tool for analysis of motor task performance. It measures electrical signals associated with muscle activation. An EMG system usually contains the following components: electrodes (electrode leads), an amplifier system, and computer software for data visualization and analysis. EMG signals are detected by electrodes and amplified. The amplifier circuit should permit adjustment of the gain for each EMG (muscle) channel. Recordings of EMG can be performed using either needle electrodes inserted into the muscle or by surface electrodes taped on the skin over the target muscle. Needle electrodes are used mostly for medical and rehabilitation purposes, extracting detailed information on muscle innervation (Göbel, 2005).

Surface EMG is a non-invasive technique for measuring muscle electrical activity that occurs during muscle contraction and relaxation cycles. It is widely used in many applications, such as: physical rehabilitation, urology (treatment of incontinence), biomechanics (sport training, motion analysis, research), ergonomics (studies in the workplace, job risk analysis, product design and certification) (Florimond, 2009). Basic (monopolar) EMG measurement uses the potential from an electrode and a ground signal as inputs to a differential amplifier. A bipolar electrode arrangement uses two active electrodes (about 2 cm apart from each other) and a differential amplification (Gath and Stalberg, 1976). Bipolar electrode arrangement allows for improved direction of signal detection with reduced cross-talk from other muscles. In such an arrangement, only the signal difference between both active electrodes is amplified, and common signal portions (from muscles located farther away) are suppressed.

There are various sources of noise in EMG signal collection, thus the muscle activation signal needs to be filtered. Typically, a low pass filter of 500 Hz is used to eliminate noise from electromagnetic drift in wires and electrodes of the system or from poor insulation. A high pass filter of 20 Hz is used to eliminate noise from poor electrode contact, variation of skin potential and motion artifacts. A notch filter of 60 Hz (frequency of power line current in the U.S.) should also be applied to control the signal contamination. Prior research has demonstrated the effectiveness of these filters for muscle signal isolation. Marras (1990) discussed the use of EMG for analysis of performance associated with the workplace activity. The technique is often used to compare the specific musculoskeletal stress associated with various work positions and postures in light and repetitive activities. Marras (1990) said the information that can be obtained from EMG analysis falls into four categories: (1) knowledge about whether the muscle was in use during an exertion; (2) muscle exertion level determined based on comparisons of levels of the processed signal under various conditions; (3) quantitative information regarding force generation of the muscle with exertions performed under restricted conditions (e.g., static or constant velocity); and (4) muscle fatigue based on analysis of the spectral components of the raw EMG signal. Related to fatigue information, as a muscle fatigues, the median frequency shifts to a lower level during a standard contraction.

In ergonomics research, comparisons of EMG levels are often made between subjects, exertion conditions and muscle groups. Consequently, normalization of the EMG signal, which provides a relative measure of muscle activity, is necessary. Normalization is usually accomplished by computing the ratio of EMG activity of a specified uptake area under an experimental condition to a reference. For example, the task EMG can be divided by the maximum EMG and represents the level of effort (in percent of max) required for a task. The resting EMG level may also be subtracted from both the task and maximum EMG. This provides an estimate of the added muscle activity necessary to perform a task independent of posture position, for example. Other methods include normalizing EMG responses to task cycle, to the maximum level of signal across an entire task, to the amplitude for a known level of force, etc. However, expressing EMG as a percentage of the isometric maximal voluntary contraction (MVC) is considered the most powerful strategy for physiologic interpretation in healthy people (Sousa & Tavares, 2012).

Marras (1990) said that the uptake area for EMG analysis needs to be carefully selected. It is also important that the area does not change between conditions. After identification of the muscle site, the electrode-muscle-interface must be prepared to ensure a high quality signal. Preparation steps typically include shaving the selected site (if necessary), cleaning the skin with alcohol, abrading the muscle area to exfoliate the skin, and applying conductive gel to the electrodes. The last two steps help reduce impedance between the muscle and the electrode at the skin surface. Impedance and noise can compromise the EMG signal and potentially lead to misinterpretation of results in a study. In general, when a muscle is at rest, the signal amplitude should be low, presenting very little or no activity. When the muscle is contracted, the amplitude of the signal should increase. The gain of an amplifier should be adjusted such that the EMG signal reveals striking changes in amplitude between exertion and resting levels.

Statistical analysis is generally used for accurate interpretation and comparison of EMG data among exertion conditions. Common techniques include the t-test and analysis of variance (ANOVA). However, if a study is interested in the influence of workplace factors on the collective behavior or multiple muscles, multivariate techniques such as multivariate analysis of variance (MANOVA) may be used to determine whether several muscles, as a group, respond in a significantly different manner to experimental conditions (Marras, 1990). As an example of EMG analysis research for facilitating human cognitive and motor performance, Sabourin and Rioux (1979) designed a study to compare the effects of "passive" muscle relaxation with "active" muscular tension on human performance of both memorization (of nonsense syllables) and simple reaction time and rotary pursuit tasks. The “passive” condition involved a subject maintaining a relatively constant low-level EMG. In the "active" condition, subjects were requested to produce different levels of muscle tension and to perform differential relaxation, i.e., reaching a specific tension level in a given muscle group and changing suddenly to the lowest possible level. Eighteen subjects were required to complete five consecutive daily sessions of training (either active or passive) or they were assigned to a no-training control condition. Subsequent to the training, participant cognitive and motor task performance was evaluated in conjunction with the assigned muscular contraction condition. Results showed that the “active” group performed significantly better than the control group and better than the “passive” group. The authors concluded that procedures for control of neuromuscular tension and reduction of tension achieved through relaxation training could be important for improving cognitive and motor abilities. This study demonstrated EMG techniques to be a reliable form of biofeedback for facilitating simultaneous task performance.

In another study of EMG analysis for motor control training, Engelhorn (1988) used EMG and kinematic variables to investigate changes in the control of movement during practice by children. He asked children in three age groups, 7, 9 and 11 yrs., to perform 60 trials of an elbow-flexion movement. Correct movements consisted of a 60° angular movement of the forearm. The analysis of biceps brachii and triceps brachii muscle EMG activity, movement displacement and timing error, and movement velocity patterns indicated changes in motor performance with practice. Results showed that all age groups improved performance with practice and exhibited a decrease in biceps EMG activity with practice. Only movement-time error and time to peak triceps muscle activity differed between the age groups. The 1l-yr.-old group significantly altered the timing of the antagonistic response to arrest movement over the practice session. This change was related to the greater information-processing ability of older children and development of appropriate movement strategies to perform the movement task successfully. Multiple advantages of the EMG analysis were identified, including: providing an indirect but superior indication of the timing of exerted force and relative changes in force produced by specific muscles during movement; providing a noninvasive window on the temporal and force parameters of muscle control during dynamic movement; and supporting observation the changes in the pattern of force production and subsequent movement attempted.

As an example of an industrial application of EMG analysis, Hagg et al. (1977) studied forearm muscular exertion during intermittent gripping tasks with three different regimes. Nine female subjects performed gripping work at 25% MVC with work-rest ratios of 1:1, 2:1 and 3:1. EMG data was collected from five forearm muscles, including three extensors and two flexors. Results revealed significant signs of fatigue in at least two of the extensor muscles across all regimes while significant signs of fatigue for the flexor muscles were only observed under the 3:1 work-rest ratio. Consequently, the authors concluded that fatigue effects were generally larger on the extensor side of the forearm. This study demonstrated EMG to be a useful and reliable tool for assessing muscle fatigue.

More recent research has also demonstrated EMG to be a useful tool in facilitating motor learning. Yiu et al. (2005) examined the effectiveness of surface EMG biofeedback for motor learning in the voice production domain. They chose surface EMG because it is commonly used as a biofeedback tool and has been tested in numerous motor domains. They conducted an experiment with two groups of participants in order to investigate whether concurrent and terminal biofeedback caused different levels of learning in a relaxed laryngeal musculature task during spoken reading. One group (concurrent feedback group) received a real-time EMG waveform display of muscle activation from bilateral thyrohyoid sites during reading trials. The other group (terminal feedback group) received a static terminal EMG waveform display about activation level from the same sites upon completion of successive trials. The study results showed no clear evidence of reliable changes in the targeted laryngeal muscle activation level across the training for either the concurrent or the terminal feedback groups. However, reliable decreases were seen in muscle activation from orofacial sites on which no-feedback was provided to participants. The decreases were equivalent across concurrent and terminal feedback groups. Although these results were unexpected by the researchers, this study supported the effectiveness of EMG as a biofeedback tool.

As another example of research using EMG techniques to provide patients with sensory feedback and promote motor learning, Casellato et al. (2013) prototyped an innovative VR apparatus that provided users with EMG-based visual-haptic biofeedback during upper limb movements (a spiral tracking task). They sought to determine whether augmented sensory feedback could induce motor control improvements in patients with primary dystonia. Participants were asked to trace a drawn spiral path presented on a virtual table by moving the end-effector of a haptic device, similar to the computerized ROCF studies conducted by Li et al. (2010, 2013). A real-time control algorithm was developed to synchronize EMG response data with haptic feedback through the device; brachioradialis EMG values were used to modify visual and haptic features at the simulated task interface. As EMG levels increased in participants, the coefficient of virtual table friction was increased and the VR display background color proportionally changed from green to red.

From recordings on dystonic and healthy subjects, results revealed that the biofeedback had a significant impact on dystonic muscular control and was correlated with the degree of local impairment. The authors concluded that the effectiveness of biofeedback paradigms was promotion of greater voluntary control of specific muscles. This study further confirmed the effectiveness of EMG to provide biofeedback.

As supported by the literature, EMG techniques provide a reliable measure of muscle activity, allowing for accurate analysis and clear understanding of motor control patterns. Such measures may be complementary in nature to subjective and performance-based evaluation of motor performance. While subjective evaluations and performance measures can be influenced by a subject's personal life experiences or psychological states, muscle activity simply provides information on how a subject reacts to a task stimulus. For this reason EMG assessment of motor performance may be more accurate and reliable than performance measures. Comparison of muscle activation levels under different motor task conditions can provide a clear picture of how task features influence muscle use, which is beneficial in terms of a deeper understanding of task performance outcomes. However, EMG systems typically have a number of components and many interconnections, making setup difficult and time consuming. Use of electrodes also usually requires tethering of subject appendages, which can restrict movement during a test and inhibit performance. A need exists for use of EMG measures in combination with performance responses for assessment of human motor control behaviors.

2.4 Cognitive Workload Assessment of Computerized Psychomotor Tests

Computerized versions of psychomotor tests have the advantages of ensuring consistent test delivery, saving clinician labor, providing objective and accurate measures of human performance and promoting ease of data analysis. However, such technologies may also pose new issues. First, clinicians and patients may need to train on the use of such advanced technologies. Test task interfaces are different from traditional physical task objects. Traditionally, subjects view test equipment in the real environment, which is familiar. With computerized tests, subjects usually view virtual test objects on a computer screen. In this case, the resolution of visual information may be different than in reality. Furthermore, virtual test information is often simplified and provides only cues necessary for accomplishing the task. Although VR technologies allow subjects to view a test environment in 3-D, such displays typically only support binocular disparity depth cues but do not cause accommodation of objects at different distances.

Second, the interaction style used in computerized tests is often different from traditional physical tasks. Traditional psychomotor tests are typically delivered in paper-and-pencil format or by using standardized instruments (i.e., an ROCF template; a set of colored blocks as part of the BD test). Subjects are able to touch and feel the instruments and directly manipulate them using their hands and fingers. However, for computerized versions of such tests, subjects manipulate virtual task objects using common computer interfaces, for example, a mouse, keyboard, or haptic device. In most cases, subjects need some training and extra instruction in order to become familiar with object manipulation using these devices.

Third, it is possible that the computerized versions of psychomotor tests might lead to subject frustration with methods of object viewing or manipulation. Some subjects might also fail to achieve proficiency in simulation use. Given these considerations, it is important to ensure that computerized versions of psychomotor tests do not introduce new forms of cognitive load relative to native test task design. Unfortunately, the majority of previous research has focused only on performance assessments with computerized psychomotor tests and few if any analyses have examined cognitive workload effects.

Many different types of measures have been developed for cognitive workload assessments. In general, several attributes are important to measure design (O'Donnell and Eggemeier, 1986). First, sensitivity implies that a measure must reveal changes in task difficulty and cognitive resource demands, and that the resolution of the measure must be such that differences are detectable for practically significant changes in workload. Second, diagnosticity implies that a measure can be used to identify causes of variation in workload, and be able to relate changes in cognitive resource demand due to loading of perceptual mechanisms or response mechanisms. Third, selectivity implies that a measure must be able to discriminate among different types of load (e.g., load due to physical effort, emotional stress, etc.) and should not be affected by physical load. Fourth, unobtrusiveness implies that a technique should not intrude in task performance and the measure should not serve as distracter to a subject's task. Fifth, reliability implies that a measure must yield repeatable results under comparable test conditions. Common measures of cognitive workload include: primary task performance, secondary task performance, physiological responses and subjective ratings. These types of measures are reviewed in the following subsections.

2.4.1 Primary and secondary task performance measure

Primary task performance measures can be used to reveal level of task difficulty. It is expected that a task with higher workload will be more difficult, resulting in degraded performance compared with a low workload task. Primary task performance is effective as an index of long periods of workload, overload and individual differences in resource use (Stanton, 2005). However, primary task performance has been found to be insensitive for revealing workload differences at low and medium difficulty levels. That is, if an increase in workload is within the attentional capacity of an operator, performance may not be degraded. Therefore, primary task performance is not a reliable measure of workload in isolation.

In order to address this issue, secondary-task techniques have been developed and are usually used in conjunction with primary performance, in order to infer the level of mental workload in a task. To use primary and secondary task measures, participants should be instructed to maintain consistent performance of the primary task while attempting to perform the secondary task when primary task demands allow. Secondary tasks must be carefully designed in order to ensure a measure of spare attentional capacity. Regarding secondary task design, Brown (1978) suggested that researchers adopt discrete stimuli that occupy the same attentional resource pools as the primary task in order to ensure the technique is measuring spare capacity and not an alternative resource pool. Secondary tasks have the advantage of being able to discriminate between tasks when no differences are observed in primary task performance. Such measures are also useful for quantifying short periods of workload, limited spare attentional capacity and even automaticity (Stanton, 2005). The primary limitation of such measures is that they can be intrusive to primary task performance, particularly when the workload of the primary task is low.

As an example of the use of secondary task measures of workload, Grant et al. (2013) proposed a method for measuring surgeon mental load using a simple secondary task in a surgical context. The primary task was a surgical training task in which participants manipulated small objects with surgical graspers. The secondary task required participants to say the word “time” each time they thought that 21 seconds had elapsed. Participants were instructed to complete the surgical task as quickly and accurately as possible and to attempt to make the intervals in the time-keeping task (secondary task) as accurate and consistent as possible. It was emphasized for participants that performance of the surgical task was more important than the time-keeping tasks. Grant et al. found this simple approach to workload assessment using the secondary task to be sensitive to primary surgical training task manipulations.

2.4.2 Physiological measures

Physiological measures of mental workload make use of various bodily response measures, including respiration and heart rate (Roscoe, 1992), heart rate variability (Jorna, 1992), electrodermal responses (Helander, 1978), eye movements and pupillary responses (Backs and Walrath, 1992), and event-related potentials (Kramer et al., 1996) as indices of mental effort.

Heart-rate variability (HRV) has been found to be a good indicator of overall demands placed on cognitive resources. Variability in heart-rate over time decreases with increased cognitive task load. Jorna (1992) provided a literature review of implementations of spectral analysis of heart rate as a tool for study of mental workload and stress in laboratory and field studies. Jorna (1992) found that such physiological measures have been used to study changes in energetical state as a function of stressors (e.g., noise, sleep deprivation, drug use) or to index workload as a function of task parameters as well as subject involvement. One research effort was aimed at validating HRV as an index of mental workload, based on the observation that a decrease in HRV corresponded with mental loading. However, Jorna (1992) pointed out that this relationship was complicated for two reasons: HRV can be influenced by several non-psychological factors, making it difficult to discriminate between physical and respiratory effects of task parameters; and the concept of "task difficulty" is ambiguous and can be dictated by many factors. Related to this, Jorna (1992) said that spectral analysis is a preferred method to obtain HRV measures because it reveals the total frequency content of cardiac signals. He said that spectral measures appear to be very sensitive to task-rest differences but less sensitive to increased difficulty levels within the same type of task. Only major changes in task structure appear to induce sizable and significant HRV effects. Jorna (1992) concluded from his review that individual subjective appraisal has an influence on both psychological and cardiovascular states. He also noted that working memory load was one of the few task parameters to reveal HRV effects and that this result corresponded with subjective evaluations of workload in working memory.

Electro-encephalography (EEG) is a measure of brain activity. Greater cognitive load causes amplitude of positive brain waves to decrease. In a flight simulation experiment, Song et al. (2011) recorded primary and secondary task workload measures along with pilot EEG during takeoff, cruise and landing activities. They found that the main effect of mental workload was significant in peak EEG amplitude and peak latency of an event-related potential (ERP; a positive wave form with a 300 ms lag relative to stimulus presentation (P300)), indicating that under higher mental workload, the ERP peak amplitude decreased and peak latency extended compared to results under lower mental workload.

Pupillography measures pupil diameter. Diameter has been found to increase as cognitive load increases (e.g., Reiner & Gelfeld, 2014). The measure is considered to be sensitive but not diagnostic in nature. Pupil size has been shown to reflect processing load or mental effort and systematic variation in pupil diameter in response to stimuli has been largely used as a measure of task-related information processing on cognitive tasks. Reiner and Gelfeld (2014) estimated mental workload through event-related fluctuations of pupil area during a task in a virtual world. They developed a non-intrusive, objective method to estimate mental workload in an immersive VR system, through analysis of frequencies of pupil fluctuations. In the VR system, participants were asked to hold a haptic stylus and move a cursor to hit a virtual cube without seeing the motion of their own hand but seeing a virtual hand synchronized with their hand movement. Changes in mental workload were tested through a number of task-repetitions involving manipulations of prior experience and the level of task predictability. Eye-tracker data was recorded for each participant continuously during the experiment. They determined rhythms in fluctuation of pupil area and developed a power spectrum density (PSD) for each participant. The ratio of low frequency (LF) to high frequency (HF) components of pupil fluctuations, and HF only components were extracted from the PSD and used to calculate mental workload. Results showed that mental workload decreased (indicated by the decrease of LF/HF ratio) with the number of task repetitions.

Blink rate is a measure of frequency and duration of eye-blinks and is measured through changes in muscle neural energy. Under high cognitive loads, humans demonstrate fewer and longer blinks. Blink rate has been found to be an accurate and reliable indicator of workload in visual perception tasks. Benedetto et al. (2011) examined the effects of in-vehicle information systems (IVIS) usage on eye blinks in a simple driving task designed by the International Organization for Standardization. Participants performed a lane change task in a driving simulator under single- and dual-task conditions. The dual-task condition was manipulated by introducing an IVIS task in the car cockpit, in which participants were required to double-click on the portion (left or right) of the screen where a target circle was located. Two difficulty levels were used: an easy one with fewer distractors (small circles), and a difficult one with more distractors. Under both difficulty levels, targets (large circles) had a diameter of 1.4 cm (distractors 0.7 cm). Blink rate, blink duration and average pupil size were measured as physiological dependent variables. Their results suggested that blink duration (BD) was a more sensitive and reliable indicator of driver visual workload, as compared with blink rate (BR). It was also revealed that the distribution of BD followed a Gaussian-like curve under normal driving conditions. Short and long blinks reflected the effects of visual workload and time on task respectively. It was found that more short blinks occurred when the IVIS was being used during driving, while more long blinks arose as greater time was spent driving.

In general, physiological measures provide for continuous monitoring of workload state and increased sensitivity in measurement. They also have the advantage of not interfering with primary task performance. Nevertheless, physiological measures of cognitive load can be easily confounded by extraneous interference (e.g., physical workload; Stanton, 2005) and influenced by individual differences. Some equipment used to collect responses may be physically obtrusive to subject behavior. In addition, data collection and analysis can be tedious.

2.4.3 Subjective measures

Subjective ratings have been identified as possibly the only index of “true” mental workload (Hart and Staveland, 1988). They are sensitive to perceived difficulty, concurrent activities, and demand for multiple resources (Hockey et al., 1989). Subjective measures have the advantage of sensitivity to changes in effort when such effort maintains primary task performance at stable levels (Hockey et a., 1989). Subjective ratings can be unidimensional and multidimensional: unidimensional measures tend to be simpler to apply and analyze, but they offer only a general workload score; multidimensional measures provide some diagnostics for identifying the sources of mental workload, but the procedures are typically more complex (Stanton, 2005). Although subjective ratings are reliable and simple, they also have limitations. They typically can only be administered post-task, which can influence reliability for long task durations. Furthermore, metacognitive limitations can cloud accurate reporting. It can also be difficult to make absolute comparisons between subjects due to internal differences in perceptions of workload (Stanton, 2005).

The subjective workload assessment technique (SWAT) is one of the most common subjective measures, which was developed by Reid and Nygren (1988). SWAT measures workload in three dimensions – time pressure, effort and task stress. Interval integer scales (from 1 to 3) are used. Subjects rank in ascending order 27 index cards with different combinations of the three dimensions, which represent workload. However, the scales do not allow for fine discrimination of different workloads. Output values of the technique are discrete and not conducive to simple parametric statistical analysis.

The NASA Task Load indeX (NASA-TLX) was developed by Hart & Staveland (1988). It also involves ranking of workload demand components (Effort, Performance, Frustration, Temporal demand, and Mental and Physical demand) in terms of perceived importance to a task. A total of 15 pair-wise comparisons of demands are used. Each demand component is rated subsequent to task performance by using a 5-in bi-polar scale with anchors of "low" and "high". Rankings and ratings are combined to compute an overall workload score. Currently, computerized versions of this technique are available.

The Modified Cooper-Harper (MCH) Scale is another subjective workload analysis technique that is particularly applicable to dual-task performance (e.g., pilot evaluation of new aircraft in terms of flight handling and stability). The MCH uses a 10-point unidimensional rating scale that yields a global rating of workload. A decision tree can be used to assist a rater in determining the most appropriate rating to assign to a system. The MCH scale has the advantage of requiring less time for workload ratings than some other techniques. It is a measure that also has great utility for field applications (Hill et al., 1992).

Hill et al. (1992) compared four subjective workload rating scales in a series of studies, including SWAT, MCH, NASA-TLX and Overall Workload (OW). The OW is literally a subject's rating of overall workload on a unidimensional scale from 0 to 100, with 0 representing very low workload and 100 representing very high workload. In five studies with three U.S. Army systems, the scales were compared with one another in terms of four dimensions: sensitivity (as measured by factor validity), operator acceptance, resource requirements, and special procedures. The authors found the NASA-TLX to have the greatest correlation with the operator workload factor and identified this technique to have the highest factor validity. Operator acceptance data showed that the NASA-TLX was liked best and was rated the best in its ability to represent workload. By analyzing resource requirements, the authors also found the NASA-TLX and SWAT to require more time for data reduction and analysis because of their multidimensional nature. However, these two multidimensional measures may also be used diagnostically to address the question of what is causing the workload reported, and results may point toward ways to alleviate excessive workload.

Given the pros and cons of each method for mental workload assessment, most human factors research references recommended use of a combination of techniques when possible (Sirevaag et al., 1993). However, it is well accepted that the NASA-TLX has advantages over other subjective MWL measures, for a number of reasons. First, the TLX provides a multidimensional evaluation of workload, with the potential to identify those components of mental workload most affected by an experimental task paradigm. Second, the TLX is more sensitive to mental workload differences than SWAT, particularly at low workload levels (Nygren, 1991). Hart and Staveland (1988) stated the procedure of the TLX is practically and statistically superior to SWAT because the independent components of the TLX provide additional diagnostic information unavailable in SWAT. Finally, and most importantly, the TLX is much easier to administer than other subjective measurement techniques. Hill et al. (1992) stated the method is more acceptable to participants and increases the likelihood of genuine responses.

3. Problem Statement

3.1 Summary of literature review

Psychomotor tests have been tested in laboratory and field studies, and some have a long history of application in clinical use. Some tests are used to assess multiple aspects of subject cognitive and motor ability. Research on these psychomotor tests has been proved of great value and provided a solid basis for various clinical or industrial applications. The progress of information technology has made it possible to computerize some widely used psychomotor tests. Computerized versions of tests provide advantages over traditional versions by providing more objective and accurate measures of performance, standardizing experiment delivery processes and reducing human labor. In order to validate the use of computerized versions of psychomotor tests and confirm their reliability, various research studies have been conducted to assess whether performance results are similar to results obtained with original tests or within a certain range of expectation.

Few studies have considered the use of other human response measures, such as muscle activation and cognitive workload, as bases for determination of reliability or validation of computerized tests. Performance of psychomotor tests may be influenced by various factors and it may not be safe to conclude that computerized versions are equivalent to physical tasks or appropriate for future testing and research on this basis, alone. Given this issue, physiological measures may provide more direct and clear indications of how subjects react to computerized psychomotor tests. For example, EMG analysis may provide a better understanding of any difference in subject performance of computerized vs. traditional psychomotor tasks. The use of physiological response measures may allow for better explanation of reliability and validity of computerized versions of tasks.

Related to this, it is possible that computerized tests may pose new types of mental workload on subjects. Lack of familiarity of digital displays and task interaction styles may cause extra workload for subjects; therefore, performance may be degraded and the reliability of study results may be reduced. Hence, it is important to ensure that computerized version of a psychomotor tests do not introduce increased workload relative to the original form of tests.

3.2 Research Expectations

Several research expectations were formulated for the present study, including:

(1) EMG measurement and cognitive workload assessment can be integrated with delivery of both computerized and physical versions of psychomotor tests.

(2) The use of physiological response measures and cognitive workload analysis are expected to complement performance assessment in terms of providing an accurate and complete evaluation of user behavior in computerized versions of psychomotor tests relative to the native versions of tests.

(3) EMG measures will be sensitive and reliable indicators of levels of psychomotor test difficulty within test type.

(4) Cognitive workload measures will be sensitive and reliable indicators of levels of psychomotor test difficulty within test type.

4. Method

An experiment was conducted to investigate the utility of EMG and cognitive workload analysis as complementary methods to performance analysis as bases for determining the validity of computerized versions of psychomotor tests relative to native or physical testing.

4.1 Tasks

The main task of the current experiment was the VR-based BD task. Administration of the BD test followed the WASI protocol (The Psychological Corporation, 1999). Designs 4-9 of the WASI booklet were used as stimuli and required participants to use four colored blocks to reconstruct the design patterns. Participants were instructed to complete each trial as quickly and accurately as possible.

The present study made some modifications based on the BD task described above. In both native and VR conditions, a specific workspace was identified in which participants were required to perform rotation of blocks for ultimate placement in an assembly grid. Participants were instructed to not perform block rotation in any other task space. When participants moved a block out of the rotation space, they were permitted only to perform transitional movements without changing the orientation of the block. This constraint was added to the original WASI protocol in order to separate the rotation phase of block movement from the position honing phase (i.e., transitional movement). This approach was necessary in order to facilitate EMG analysis of specific muscle use during task performance. Definition of the rotation space ensured that muscle responses while reorienting blocks could be analyzed and compared separately from responses collected during block displacement.

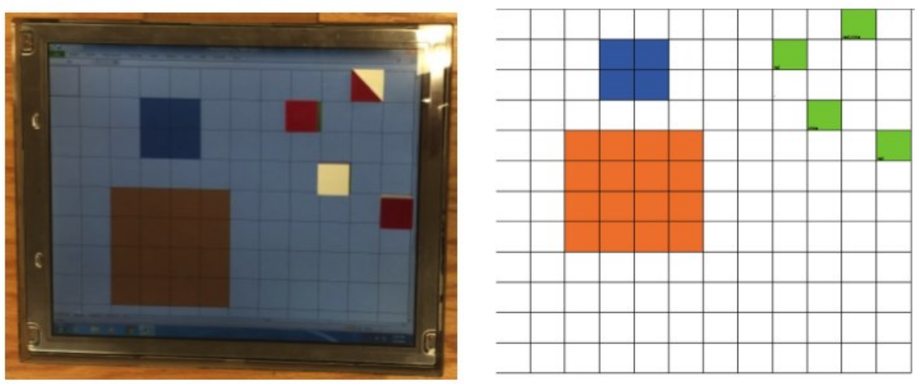
Each participant assigned to perform the VR-BD test was required to also complete a dice-manipulation training task before the formal experiment test. This program was designed to familiarize participants with use of the haptic stylus control, including the block-shaped interface to manipulate virtual blocks. Participants were trained in the task for proficiency with the VR system. In every trial of the dice-training task, a single virtual die was presented near the center of a display on a virtual work surface. A two-dimensional stimulus design was presented at the top of the display area. The participants were instructed to move the die as quickly and accurately as possible to the target square near the right bottom of the work surface, with the top surface of the die matching exactly the stimulus. This training program was developed in a prior study (Clamann et al., 2013) and proved efficient in familiarizing participants with the haptic device in a series of 40 trials. Pilot tests of the training program as part of this research also revealed no significant fatigue effect due to the device training; therefore, only participants under the VR condition were required to complete this part of the experiment protocol.

4.2 Apparatus

4.2.1 Native BD task

A custom workstation with a flat-screen monitor was mounted in a wood tabletop and was used as a basis for performing the native BD task (see Figure 1). An Excel file (see Figure 1) was presented on the display with highlighted cells indicating starting block positions, a destination grid, and block rotation area for participants. The green cells in Figure 1 show where the blocks were initially located. The orange area served as the rotation area. The blue grid represented the target area where participants were asked to position blocks according to the designated design pattern. All block positions presented on the display were consistent with those in the VR version of the task.

In order to precisely observe participant performance, a Logitech® HD Webcam C525 was mounted on the ceiling of the lab on a rack directly above the monitor in the table. Prior to each trial, participants were asked to sit erect and a goniometer was used to help them flex their neck by 20 degrees, which is a comfortable posture for performing the task. While performing the task, the muscle activation of participants was recorded.



(a) Native BD workstation **(b) Location indicators in Excel**

Figure 4.1 Custom workstation for native BD test

4.2.2 Computerized BD task

The VR interface for the BD task (see Figure 2) was presented on a PC integrated with a stereoscopic display using a NVIDIA® 3D Vision™ Kit, including 3D goggles and an emitter. Stereoscopic rendering of the task simulation was supported by an OpenGL quad-buffered stereo, high-performance video card (NVIDIA® Quadro™). The SensAble Technologies PHANTOM Omni® Haptic Device was used as the haptic control platform and integrated with the PC. The Omni includes a boom-mounted control that supports 6 DOF movement and 3 DOF force feedback. The interface automatically recorded participant performance data.

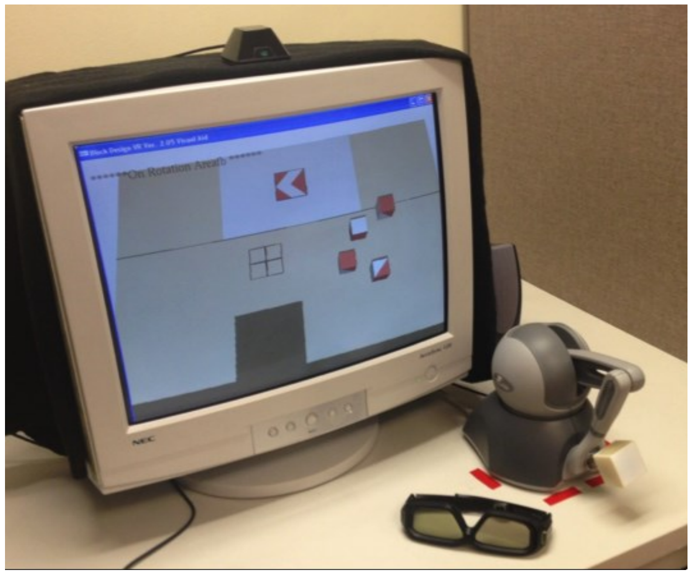


Figure 4.2 Virtual reality workstation for Block Design test

The VR-BD simulation task included a virtual tabletop divided into two parts, including a display area and a work area (see Figure 3). The display area presented the stimulus design patterns to be replicated by a participant. The work area was used for arranging the virtual blocks. The work area and blocks were presented at approximately 70% of actual size to allow the design pattern and workspace to be viewed on a 21-inch stereo monitor. The view angle of the work area was set to 70 degrees from horizontal for a condition comparable to participant view angle of the flat-screen monitor in the native task condition. Similar to the native version, all BDs were constructed with the aid of a target grid, which appeared as a 2x2 collection of squares in the work area. A shaded area was presented at the bottom of the display to identify for participants the block rotation area in the virtual task.

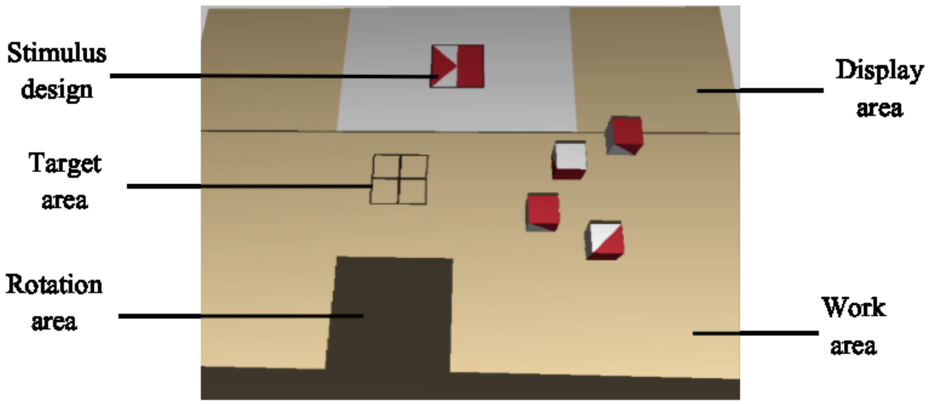


Figure 4.3 VR-based BD test simulation

During the BD testing, the Omni haptic device with customized block control interface (as shown in Figure 2) was used to manipulate a cursor (small blue orb) appearing on the VR display. Virtual blocks could be grasped by touching the cursor against them and pressing the button on the block-shaped control. A block could then be lifted from the virtual table surface and rotated about any axis using the haptic control. Haptic features were included in the simulation to represent the blocks and the table as solid objects.

4.2.3 EMG system

Muscle activation data was recorded using the BIOPAC® MP 150 system. Disposable Ag/AgCl electrodes (with circular contacts of 11 mm diameter and square cloth backing) were used. For each participant, eight electrodes were mounted across 4 muscle positions, with 2 electrodes placed at 2 cm apart for each muscle. One electrode was mounted at the elbow as a ground electrode. These electrodes were then connected to the amplifier system using leads with standard snap connections. Four amplifier modules were cascaded together, providing 4 channels of muscle output. The gain was set to 2000. The amplifier system was connected to a computer using a 2-meter Ethernet crossover cable (CBLETH2). Raw data was recorded with the "AcqKnoweldge 4.2" software.

4.3 Participants

A total of 24 right-handed participants were recruited from the North Carolina State University student population as well as the off-campus (community) population. The age range of participants was 19-46 years old (Mean = 26.4; S.D. = 6.5). All participants were required to have 20/20 or corrected vision (if corrected, they were asked to wear the glasses or lens while performing the experiment tasks) and to have neither current nor chronic wrist disorders. All sample inclusion criteria were confirmed with each person in a screening step before scheduling experimental sessions. All participants signed an informed consent form prior to experiment and were compensated by cash at a rate of $15 per hour.

4.4 Independent variables

This study aimed to investigate the equivalence of the computerized version of the BD test with the native task in terms of performance, user muscle activation and cognitive workload. Therefore, only one independent variable was identified for the study; that is, the test condition including VR vs. native testing. These conditions served as two levels of the independent variable.

4.5 Dependent variables

Three types of dependent variables were collected during the experiment, including the EMG responses, BD trial completion time, and NASA-TLX workload ratings.

(1) The muscle activation responses were recorded for the flexor pollicis bravis, pronator teres, extensor carpi radialis, and flexor carpi ulnaris during each test trial. All EMG responses were divided into rotation and honing phase data. The rotation phase data represented any activation that occurred while a participant held a block in the rotation space. Honing phase data represented muscle activations when a block exited the rotation area and was moved to the target grid area.

(2) The VR-BD program or experimenters recorded time-to-task completion (TTC) for each design during a test trial. The total time spent in completing an entire trial was used as a response for further analysis. Since all block Designs 4-9 of the WASI were completed in each trial and any design effect on reconstruction time was not of interest in comparison of the VR vs. native versions of the task, examining total trial completion time was a way of simplifying the data analysis.

(3) The workload that participants experienced under both conditions was assessed using the NASA-TLX. Rankings of demand components were completed prior to administration of the BD trials and ratings for the six dimensions were collected at the end of each trial. The demand rankings and ratings were integrated to determine the total TLX score.

4.6 Experiment design

This study followed a between-subjects experiment design in which each participant was randomly assigned to either one of the BD task conditions, namely the native or VR setting.

In order to determine whether the different test conditions resulted in comparable responses, a within-subject design would be effective for mitigating the confounding effect of individual differences in performance. However, with the BD task, it is possible that cognitive processing and physical movement learning effects may occur from one condition to another. Therefore, a between-subjects experiment design was selected for the present study to avoid this learning effect. With this experiment design, individual differences contribute to the observed difference in responses between the two test conditions, and must be taken into account when comparing test conditions. For this reason, participants were provided with rigorous task training in advance of test trials. Furthermore, a participant term was included in all statistical models in order to separate any response variance attributable to individual differences from response variance due to the test manipulation. Among the two groups of participants, those persons assigned to the native test condition performed the conventional BD task using the custom workstation; for those assigned to the VR test condition, the task was performed using the VR-BD simulation with the custom block-shaped haptic control device, as described above.

4.7 Procedures

Participants were initially provided with a brief introduction to the lab environment and experiment. After signing an informed consent form, EMG electrodes were placed on hand and forearm of participants. Activation of all four identified forearm muscles was measured in the study. Only right forearm muscles were monitored since all tasks were completed with the right hand.

To establish a baseline for comparison of muscle activity levels under various conditions, maximum voluntary contractions (MVC) were recorded for each muscle group. The MVC test tasks were designed to explore the maximum force individual participants could generate with specific muscle. Muscle activation levels occurring during BD trials were ultimately normalized in terms of MVCs on an individual basis in order to facilitate comparison of the two task conditions.

Each participant was randomly assigned to one of the test conditions (i.e., VR vs. native). Participants assigned to VR condition performed the dice-manipulation training with the haptic device, as mentioned above. A total of 40 trials were completed to cause fluent control of the device. Participants were allowed a 5-minute rest following the dice training. Participants assigned to the native task condition started a BD test session directly following MVC data collection. Both groups completed three BD task sessions with each session involving 6 trials (i.e., WASI Designs 4-9). Each session took 5-7 minutes for the VR group and 1.5-2 min for the native group (with gross time differences being primarily attributable to apparatus reset and condition setup).

At the end of each BD task session the standardized NASA-TLX workload form (Hart & Staveland, 1988) was administered to participants. In addition to pairwise comparisons of the six workload demand components, participants completed ratings for each dimension based on an experimental session, and not any specific BD trial.

The entire experiment was conducted in a separate laboratory room without any distractions to participants. Each participant took about 2 hrs. to complete the study.

4.8 Research hypotheses

If the design of the computerized version of the BD task simulation is comparable to the native task in terms of visual appearance and physical interaction, then there should be no differences among participant performance, EMG and cognitive workload responses across task types. Consequently, several hypotheses (H) were formulated for this research based on consideration of the independent and dependent variables:

H1: Based on the prior research and design of the VR-BD task simulation, BD test performance was expected to be comparable under native and VR conditions, as reflected in the task completion time response.

H2: Based on the similar performance results reported by Clamann et al. (2012) for VR and native versions of the BD task, forearm muscle EMG responses were expected to be comparable under native and VR conditions of the task, as reflected in EMG signal amplitudes.

H3: Due to the use of different forearm muscles in block rotation and translation, muscle activation levels were expected to be significantly different among the rotation versus honing phases of the BD task. Furthermore, since rotation of blocks required finer hand movement control than block displacement, it was expected that EMG responses would be larger in the rotation phase.

H4: Due to differences in the movement functions of the various forearm muscles, different activation levels were expected among the four muscles monitored in the experiment.

H5: Based on the similar performance results reported by Clamann et al. (2012) for VR and native versions of the BD task, in this study, cognitive workload responses were also expected to be comparable in BD task completion under native and VR conditions, as reflected in NASA-TLX scores.

5. Data analysis

5.1 Diagnostics and transformation

Diagnostics were conducted on all response measures in advance of any parametric inferential statistics. Diagnostics revealed normality violations for the EMG responses, time-to-task completion and the workload measure. For some response measures, the constant variance assumption was also violated. Therefore, appropriate transformations were attempted on all responses; however, they were not successful in several cases. Therefore the EMG responses and time-to-task completion were ranked and submitted to parametric procedures to perform non-parametric analysis. Workload responses were not transformed and analyzed using non-parametric procedure. Untransformed response data were also analyzed using parametric procedures. Results for the EMG responses were similar to those from non-parametric analysis, indicating that the parametric analysis was not sensitive to the identified assumption violations. Therefore, for EMG responses, only the results from the parametric analysis are reported in this document.

5.2 EMG responses

The EMG data was collected with a sampling rate of 1000 Hz. All responses were preprocessed with a fourth order Butterworth filter (high-pass filter set as 20 Hz and low-pass filter set as 450 Hz) and notch filtered at 60 Hz. The screened data was then rectified and averaged across every 51 data points for smoothing purposes. Mean values of EMG responses for each trial were calculated to represent overall muscle activation level under the different conditions. Statistical analysis of the EMG responses included the following steps: (1) Descriptive statistics were determined for both the rotation and honing phases of the BD within the native and VR test conditions.

(2) A multivariate analysis of variance (MANOVA) was conducted on the EMG responses to identify any effect of movement phase or trial number. Trial number did not prove to be significant, suggesting no participant fatigue. The trial number term was, consequently, excluded from additional model-based analyses. The MANOVA yielded Wilk's lambda as well as exact F test results. All degree of freedoms reported in the document were taken from SAS output.

(3) Subsequently, a MANOVA was used to identify differences between the test conditions (VR vs. native) within phase of movement.

(4) Finally, a series of ANOVAs were conducted to identify any effects of muscle type and test condition (VR vs. native task).

5.3 Task completion time

Task completion time was recorded for each participant in all the three BD test sessions. Analysis of the response measure included the following steps: (1) determination of descriptive statistics (mean and standard deviation) for the two test conditions (VR and native task); and (2) application of a repeated measures ANOVA to identify any effect of test condition with trial number as a covariate. In the statistical model, test condition was considered as a fixed effect and participant was considered as a random block effect.

5.4 Workload (NASA-TLX)

A two-sample Wilcoxon rank sum test was used to compare the overall NASA-TLX scores across the two test conditions (VR vs. native). Furthermore, analysis of ratings on each dimension of the NASA-TLX were performed in order to identify those components of cognitive workload that were most influenced by the test condition manipulations. This analysis was expected to provide a deeper understanding of the causes for any potential difference in workload between the two test conditions.

5.5 Correlation among response measures

Last but not least, correlation analyses were performed on the various response measures (EMG signals, task completion time, and workload ratings) within movement phase (honing vs. rotation) and test condition (native vs. VR) in order to reveal any potential linear relations among the measures. Spearman correlation analysis was used due to violations of the normality assumption for several response measures.

6. Results

6.1. Muscle activation

Descriptive statistics of muscle activation are summarized in Table 6.1 with grouping by test condition (native vs. VR) and phase of movement (rotation vs. honing). On average, muscle activation was greater during the rotation phase, as compared to honing, under either test condition. With the exception of the flexor pollicis brevis, all muscles showed similar levels of response between the native and VR test conditions.

Table 6.1 Descriptive statistics summary for EMG responses (% of MVC)

	CH	Honing		Rotation	
		Native	VR	Native	VR
Mean (SD)	1	4.97 (6.24)	2.23 (1.53)	11.06 (9.78)	7.25 (3.82)
	2	0.83 (0.53)	0.90 (0.59)	1.92 (1.51)	2.57 (1.19)
	3	0.71 (0.55)	0.90 (0.71)	1.79 (0.93)	2.81 (1.82)
	4	1.62 (0.89)	1.64 (1.05)	4.04 (1.62)	4.58 (1.94)

* CH1 - Flexor pollicis brevis CH2 – Pronator Teres

CH3 – Extensor carpi radialis CH4 – Flexor carpi ulnaris

To further analyze the EMG responses, a MANOVA test was conducted with the objective of identifying any significant effects of test condition and phase of movement. Although each participant performed three trials of the BD task under the native or VR test condition, the MANOVA revealed no significant trial effect (see Table 6.2). These results indicated no significant occurrence of participant fatigue during the experiment trials. Despite the absence of a trial effect, significant differences were detected between the rotation and honing phases of movement for the EMG responses (see Table 6.2). Based on this result, response data for the two phases of movement were separated for further analysis.

Table 6.2 MANOVA results to test trial and phase effect

	Wilks' Lambda	F value	P value
Trial effect	0.9772	$F(8,226) = 0.33$	0.9550
Phase effect	0.2520	$F(4,113) = 83.84$	< 0.0001*

For each phase of movement, a MANOVA was applied to the EMG responses in order to detect any significant effect of test condition. As shown in Table 6.3, no significant differences were detected between the two conditions. That is, the two versions of test yielded similar EMG responses within rotation and honing movement.

In order to identify any potential differences in activation levels among muscles within each phase of movement, a two-way ANOVA was applied with muscle type and test condition group as predictor variables. The ANOVA results are summarized in Table 6.4. In the statistical model, the test condition was treated as a between-subject factor while the muscle type was a within-subject factor. Significant differences were revealed among the four muscles under both phases of movement. Consistent with the previous MANOVA results, there was no main effect of test condition on the EMG responses in the honing phase or in the rotation phase. There was, however, a significant effect of interaction between the test condition and muscle type.

Table 6.3 MANOVA results on test condition effect

Honing Phase	**Wilks' Lambda**	**F value**	**P value**
	0.8302	$F(4,19) = 0.97$	0.4461
Rotation Phase	**Wilks' Lambda**	**F value**	**P value**
	0.7862	$F(4,19) = 1.29$	0.3085

Table 6.4 Two-way ANOVA results on test condition and muscle effects within phase of movement

		ANOVA Statistics
Honing Phase	Condition (C)	$F(1,22) = 2.43$, $p = 0.1330$, $1\text{-}\beta = 0.3203$
	Muscle (M)	$F(3,258) = 24.02$, $\mathbf{p < 0.0001^*}$, $1\text{-}\beta = 0.9461$
	C * M	$F(3,258) = 7.08$, $\mathbf{p = 0.0001^*}$
Rotation Phase	C	$F(1,22) = 0.39$, $p = 0.5365$, $1\text{-}\beta = 0.0923$
	M	$F(3,258) = 52.85$, $\mathbf{p < 0.0001^*}$, $1\text{-}\beta = 0.9998$
	C * M	$F(3,258) = 6.53$, $\mathbf{p = 0.0003^*}$

As an extension of the ANOVA results, post-hoc analysis was conducted on the EMG responses for each muscle type within the specific phases of movement by using a Tukey-Kramer test method. Figures 6.1 presents bar charts of the EMG responses across muscles. Those responses values appearing with the same grouping letter (e.g., “A”, “B”) are not significantly different, one from the other. For the pronator teres (CH2), flexor carpi ulnaris (CH3) and extensor carpi radialis (CH4), native and VR test condition responses appeared comparable. For the flexor pollicis brevis (CH1), EMG responses were significantly higher under the native condition as compared to the VR condition ($p = 0.0003$ under honing phase; $p = 0.0045$ under rotation phase). In addition, CH1 (pollicis brevis) showed significantly higher activation levels than any of the other three muscles within either phase of movement ($p < 0.0001$).

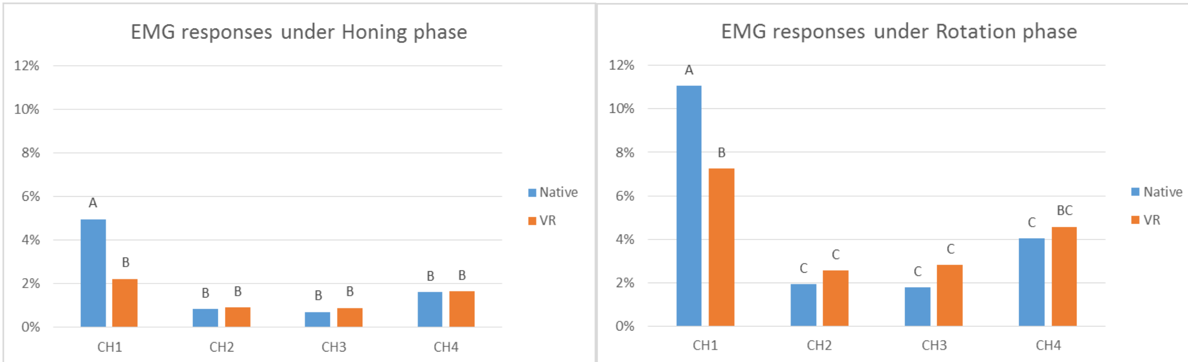


Figure 6.1 Results of post-hoc analysis on EMG responses

6.2 Task completion time

The means and standard deviations of time-to-task completion are summarized in Table 6.6 and are grouped by condition and trial. A decreasing trend was observed across trials under both test conditions. In addition, participants assigned to the VR condition took longer to complete the same task trials than those assigned to the native task condition.

Table 6.5 Mean (SD) summary for BD time-to-task completion time (sec.)

	Trial 1	**Trial 2**	**Trial 3**
VR	416.09 (212.88)	350.07 (155.95)	302.59 (142.73)
Native	131.45 (45.56)	109.75 (43.98)	95.76 (29.12)

In order to further analyze the task time data, a generalized ANCOVA test was conducted, in which trial was used as a covariate to model an overall linear growth trend. Results are presented in Table 6.6. Results showed significant effect of test condition and trial number on the time spent to complete a trial of the BD task. The descriptive statistics revealed that longer time was spent to complete a trial of the BD task under the VR condition as compared to under the native task condition. In addition, participants showed a significant decrease in completion time in later trials. The non-parametric ANCOVA did not reveal significance of the interaction effect. However, the untransformed data revealed the VR group to achieve a greater decrease in task completion time, as shown in Figure 6.2.

Table 6.6 Results of ANCOVA test for time-to-task completion

	ANCOVA Statistics
C	F(1,22) = 49.99, p < 0.0001*
Trial (T)	F(1,46) = 74.01, p < 0.0001*
C*T	F(1,46) = 0.12, p = 0.7260

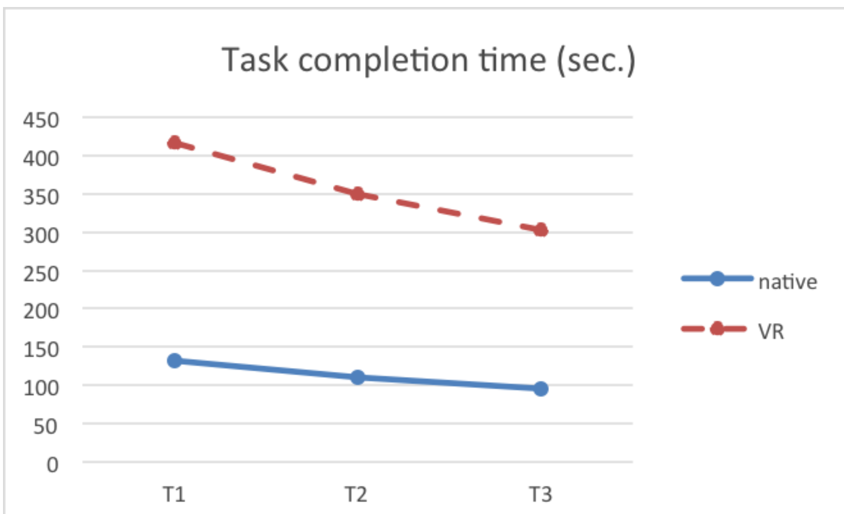


Figure 6.2 Task completion time of native and VR conditions across trials

6.3 Subjective measure of workload

The descriptive statistics for the overall NASA-TLX workload score and ratings of the six sub-dimensions are presented in Table 6.7. On average, participants under the VR test condition provided higher TLX scores, indicating higher perceived workload during performance of the BD task than the native test condition. This outcome was also reflected in relatively higher ratings for the sub-dimensions of physical, temporal, effort and frustration given by VR participants.

Table 6.7 Mean (SD) summary of TLX overall score and ratings of sub-dimensions

	TLX score	Mental	Physical	Temporal	Performance	Effort	Frustration
Native	44.26 (15.75)	49.39 (28.84)	31.97 (22.87)	45.83 (20.49)	29.64 (13.86)	48.36 (26.15)	31.63 (31.74)
VR	61.78 (7.56)	48.50 (23.00)	43.17 (30.66)	63.52 (25.42)	37.43 (17.93)	70.97 (21.24)	68.24 (25.52)

A series of two-sample Wilcoxon rank sum tests were conducted on the overall TLX score and the sub-dimension ratings. Results are summarized in Table 6.8. Participants assigned to the VR condition produced significantly higher ratings of temporal demand, effort and frustration. These three dimensions were the main reasons contributing to the significantly higher overall TLX score for the VR users. Meanwhile, no significant difference was detected in physical or mental workload, as well as participant ratings of their own performance. These findings suggested that, across the two test conditions, BD task workload was comparable in certain aspects and that participants rated their own performance at comparable levels.

Table 6.8 Results of two-sample t-tests on TLX score and sub-dimension ratings

	Wilcoxon statistics	**P-value**
TLX score	204.00	0.0006*
Mental demand	149.50	0.4947
Physical demand	163.50	0.2255
Temporal demand	180.50	0.0403*
Performance	168.50	0.1547
Effort	183.50	0.0288*
Frustration	195.00	0.0041*

6.4 Correlation analyses on response measures

Spearman correlation analyses were conducted to assess the strength of any monotone associations among the three types of response measures, within test condition as well as phase of movement. Correlation statistics for the time-to-task completion (TTC) and EMG responses are presented in Table 6.9. The performance responses for each trial were analyzed for association with the muscle activation levels. Under the native test condition, activation level of CH1 (flexor pollicis brevis) and CH4 (flexor carpi ulnaris) positively correlated with TTC in the honing phase of the BD task; activation level of CH1, CH3 (extensor carpi radialis) and CH4 positively correlated with TTC in the rotation phase of the task. Under VR condition, the activation level of CH1 positively correlated with TTC in honing and activation level of CH3 positively correlated with TTC in rotation.

Table 6.9 Correlation of time-to-task completion and EMG

				CH1	CH2	CH3	CH4
Native	Honing	TTC	Spearman's rho	0.4193	-0.0018	0.1279	0.4064
			P-value	**0.0109***	0.9917	0.4572	**0.0139***
	Rotation	TTC	Spearman's rho	0.425	-0.019	0.3459	0.6476
			P-value	**0.0098***	0.9122	**0.0388***	**<0.0001***
VR				**CH1**	**CH2**	**CH3**	**CH4**
	Honing	TTC	Spearman's rho	-0.2488	-0.018	0.2988	-0.2944
			P-value	**0.0371***	0.9169	0.0766	0.0761
	Rotation	TTC	Spearman's rho	-0.1712	0.199	0.4654	-0.0615
			P-value	0.3182	0.2447	**0.0042***	0.7215

Correlation statistics on the NASA-TLX component ratings and EMG responses for all channels of muscle activity, within test condition and phase of movement, are presented in Tables 6.10 and 6.11. Under the native test condition, overall TLX scores and physical demand ratings were positively correlated with CH3 activation level during honing; frustration ratings were positively correlated with CH1 activation level in the same phase of movement. In the rotation phase, overall TLX scores and physical and temporal demand ratings were positively correlated with CH3 activation level; frustration ratings were positively correlated with CH1, CH3 and CH4 activation levels.

Under the VR test condition, overall TLX scores were positively correlated CH2 (pronator teres), CH3 and CH4 activation levels during honing; mental demand, physical demand and effort ratings were positively correlated with CH2 activation level in the same phase of movement; and temporal demand ratings were positively correlated with CH3 activation level in honing. In rotation phase, overall TLX scores were positively correlated with CH2 and CH3 activation levels; mental demand and physical demand ratings were positively correlated with CH2 activation level; and temporal demand ratings were positively correlated with CH3 activation level.

Correlation statistics on the NASA-TLX component ratings and TTC, within test condition, are presented in Table 6.12. Results revealed that under the native test condition, mental demand and frustration ratings were positively correlated with TTCs. Under the VR test condition, performance ratings were also positively correlated with TTCs for Trials 2 and 3.

Table 6.10 Correlation of NASA-TLX scores and EMG under native condition

			CH1	CH2	CH3	CH4
Honing	**Overall TLX**	Spearman's rho	0.1201	-0.0411	0.4237	0.1782
		P-value	0.4855	0.8121	**0.0100***	0.2985
	Mental	Spearman's rho	0.2743	-0.0345	0.322	0.2056
		P-value	0.1055	0.8415	0.0555	0.2289
	Physical	Spearman's rho	0.0325	0.1859	0.5538	0.1983
		P-value	0.8506	0.2777	**0.0005***	0.2463
	Temporal	Spearman's rho	0.1121	0.0027	0.3278	0.0217
		P-value	0.5150	0.9875	0.0509	0.8999
	Performance	Spearman's rho	-0.0302	-0.0318	0.2548	0.1743
		P-value	0.8612	0.8541	0.1336	0.3093
	Effort	Spearman's rho	0.1944	-0.2114	0.1874	0.0303
		P-value	0.2560	0.2157	0.2738	0.8609
	Frustration	Spearman's rho	0.3377	0.0186	0.3168	0.3176
		P-value	**0.0440***	0.9143	0.0598	0.0591
Rotation			**CH1**	**CH2**	**CH3**	**CH4**
	Overall TLX	Spearman's rho	0.2215	-0.1239	0.5352	0.2742
		P-value	0.1941	0.4714	**0.0008***	0.1056
	Mental	Spearman's rho	0.324	-0.0675	0.4438	0.2689
		P-value	0.0539	0.6956	0.0067*	0.1128
	Physical	Spearman's rho	-0.0868	-0.0651	0.3586	-0.0310
		P-value	0.6149	0.7062	**0.0317***	0.8576
	Temporal	Spearman's rho	0.0694	0.0256	0.3410	-0.0027
		P-value	0.6873	0.8822	**0.0418***	0.9875
	Performance	Spearman's rho	0.0744	-0.2626	0.1627	0.2293
		P-value	0.6665	0.1218	0.3432	0.1786
	Effort	Spearman's rho	0.2188	-0.2720	0.2844	0.0147
		P-value	0.1998	0.1086	0.0928	0.9320
	Frustration	Spearman's rho	0.3865	-0.0682	0.4578	0.3819
		P-value	**0.0199***	0.6928	**0.0050***	**0.0215***

Table 6.11 Correlation of NASA-TLX scores and EMG under VR condition

			CH1	CH2	CH3	CH4
Honing	Overall TLX	Spearman's rho	0.0914	0.4717	0.5012	0.3509
		P-value	0.596	**0.0037***	**0.0018***	**0.0359***
	Mental	Spearman's rho	0.2781	-0.3377	-0.0271	0.1108
		P-value	0.1006	**0.0440***	0.8753	0.5201
	Physical	Spearman's rho	0.0248	0.4326	0.2894	0.2002
		P-value	0.8857	**0.0084***	0.0869	0.2417
	Temporal	Spearman's rho	0.3145	0.306	0.6026	0.2394
		P-value	0.0618	0.0696	**0.0001***	0.1597
	Performance	Spearman's rho	-0.0875	-0.1325	0.1402	-0.2409
		P-value	0.6117	0.4412	0.4147	0.1570
	Effort	Spearman's rho	0.0221	0.3414	-0.0826	0.3271
		P-value	0.8981	**0.0416***	0.6318	0.0515
	Frustration	Spearman's rho	-0.0194	0.1108	0.0008	0.0976
		P-value	0.9107	0.5201	0.9964	0.5712
Rotation			**CH1**	**CH2**	**CH3**	**CH4**
	Overall TLX	Spearman's rho	-0.0318	0.4965	0.3532	0.3137
		P-value	0.8541	**0.0021***	**0.0346***	0.0624
	Mental	Spearman's rho	0.2448	-0.4895	-0.0984	0.0573
		P-value	0.1502	**0.0024***	0.5681	0.7398
	Physical	Spearman's rho	0.0081	0.5098	0.227	0.1924
		P-value	0.9624	**0.0015***	0.1831	0.2608
	Temporal	Spearman's rho	0.2099	0.2301	0.4717	0.1851
		P-value	0.2191	0.1771	**0.0037***	0.2797
	Performance	Spearman's rho	0.1441	0.1038	0.3261	0.0891
		P-value	0.4018	0.5469	0.0523	0.6054
	Effort	Spearman's rho	-0.1622	0.2747	-0.2662	0.2588
		P-value	0.3447	0.1049	0.1166	0.1275
	Frustration	Spearman's rho	-0.0496	0.1069	0.0217	-0.0124
		P-value	0.7740	0.5349	0.9001	0.9428

Table 6.12 Correlation of NASA-TLX scores and time-to-task completion

			TTC1	TTC2	TTC3
Native	Overall TLX	Spearman's rho	0.5105	0.4336	0.4895
		P-value	0.0899	0.1591	0.1063
	Mental	Spearman's rho	0.6760	0.5990	0.5849
		P-value	**0.0158***	**0.0396***	**0.0457***
	Physical	Spearman's rho	0.0769	0.035	0.1259
		P-value	0.8122	0.9141	0.6967
	Temporal	Spearman's rho	0.0140	0.0876	0.0245
		P-value	0.9655	0.7867	0.9397
	Performance	Spearman's rho	0.3566	0.2098	0.2448
		P-value	0.2551	0.5128	0.4433
	Effort	Spearman's rho	0.0455	0.2032	0.2592
		P-value	0.8883	0.5266	0.4159
	Frustration	Spearman's rho	0.7832	0.7483	0.7762
		P-value	**0.0026***	**0.0051***	**0.0030***
VR			**TTC1**	**TTC2**	**TTC3**
	Overall TLX	Spearman's rho	-0.1888	0.1538	-0.2727
		P-value	0.5567	0.6331	0.3911
	Mental	Spearman's rho	-0.2448	-0.2867	-0.2028
		P-value	0.4433	0.3663	0.5273
	Physical	Spearman's rho	0.0946	0.1751	-0.0525
		P-value	0.7700	0.5862	0.8712
	Temporal	Spearman's rho	-0.1678	0.0629	-0.1888
		P-value	0.6021	0.8459	0.5567
	Performance	Spearman's rho	0.5245	0.6154	0.6783
		P-value	0.0800	**0.0332***	**0.0153***
	Effort	Spearman's rho	-0.3643	-0.3783	-0.5569
		P-value	0.2444	0.2253	0.0600
	Frustration	Spearman's rho	-0.5245	-0.1608	-0.4615
		P-value	0.0800	0.6175	0.1309

7. Discussion

The first hypothesis (H1) posited comparable performance among the native and VR-BD tasks. The results on task completion time did not support this hypothesis. Participants under the VR condition took significantly longer to complete the test task, which is consistent with the results of Pontonnier et al. (2014). The gap in performance time between the native and VR conditions was mainly generated by intrinsic differences between test conditions. Although the VR condition presented a customized haptic control, resembling the shape of physical block, the control still proved more challenging to use in accurately performing the task as fast as under the native condition. Furthermore, the absence of visual accommodation in the immersive “3D” virtual environment made it more challenging for participants to estimate distances from a current cursor position to a target grid location. Even with haptic features presented to simulate the touch and feel of virtual blocks as solid objects, it took longer to perceive the objects through the cursor than perception of physical blocks via direct handling.

Although extensive dice-manipulation training was provided prior to the formal experiment to help participants become familiar with the apparatus, it is possible that some participants still experienced difficulty and confusion when manipulating the virtual block in the 3D simulation. Pontonnier et al. (2014) also pointed out that a lack of familiarity of participants with their VE contributed to a difference in task completion time from real environment (RE) performance.

In addition, the native BD task had several physical advantages over the VR-BD in the present study. For example, when picking up a block, participants assigned to the native task could easily move their hands close to the physical block and only needed to grasp the block by simple finger movements. However, under the VR condition, it took more time to accurately position the cursor at a virtual block. When approaching a block, participants had to first ensure that they reached the block location with the cursor, touch the cursor to the block, and then press a button embedded in the block shaped haptic control in order to pick up the virtual block. Furthermore, under the native test condition, previously positioned blocks were "moveable" when inserting a new block into the target assembly grid. Poorly positioned blocks could be pushed slightly towards a better position when a new block was being placed. Opposite to this, the VR-BD simulation presented all positioned blocks as fixed when the cursor was "grasping" a new block. In this way, participants assigned to the VR condition had to relocate poorly positioned blocks to make room for a new block, which caused longer performance times as well as increased task frustration. Related to the present findings, in the study by Hu et al. (2007), subjects performed a simplified version of a typical hand-drilling or riveting operation in both real and virtual environments. Their data analysis showed that subjects required significantly greater time to complete tasks under the VE condition. However, a fairly good correlation of body part discomfort ratings under the VE and RE conditions revealed predictive capability of the VE simulation from an ergonomics assessment perspective. The between-subject design of the present study did not allow a correlation analysis to be conducted on the same response measures under the different test conditions (VR vs. native). Such an analysis in future study might make the relationship between performance under VR and native test conditions clearer.

The second hypothesis of this study posited that EMG responses would be comparable in completion of the BD task under native and VR conditions. This hypothesis was supported by the results. No significant differences were found in EMG responses for both the honing and rotation phases, indicating a comparable magnitude of muscle activation in task performance when using the VR workstation vs. native task setup. This suggested that participants used their forearm muscles in a similar way when interacting with the physical blocks in the native BD task and the virtual blocks in the VR-BD task. This result is important from a rehabilitation perspective as correspondence of muscle activation level across the VR and native BD tasks would be more critical to motor skill development than the resulting time-to-task completion.

Whitman et al. (2004) analyzed participant performance of lifting tasks in both virtual and real environments but, unlike the present study, they did not find comparable physiological responses among the environments. They used a head-mounted display to present the virtual environment and a lumbar motion monitor and motion capture system to measure performance data. The lifting task was to move three boxes (virtual or real) from a table of 15 or 38 inches to another table of 15 inches. The data sets for maximum lateral velocity, maximum sagittal velocity, maximum twist velocity, maximum lateral acceleration, maximum sagittal acceleration and maximum twist acceleration differed significantly between the virtual and real environments. Whitman et al. (2004) concluded that performance in virtual and real environments was not comparative. The fidelity of their virtual environment was identified as a cause for the discrepancy (Whitman et al., 2004). In the present study, the visual display of the VR-BD was of a high fidelity and the novel haptic control interface modeled the shape and size of blocks handled in the real task. These features likely contributed to the comparativeness of behavior in the VR and native version of the BD task. In addition, Whitman et al. measured and evaluated different types of physiological responses than those recorded in the present study. It is possible that the degree of sensitivity to test conditions may vary among different measurements. It would be valuable to use other forms of measurement, such as fine motion tracking, in order to extend the present study.

During a simulated simple assembly task, Pontonnier et al. (2014) collected EMG responses on five muscle, including erector spinae (back extensor), deltoideus medialis (shoulder abductor), biceps brachii (forearm supinator and elbow flexor), triceps long head (elbow extensor and shoulder stabilizer), and flexor carpi ulnaris (wrist flexor and adductor). They found significant differences in average muscle activities (AMAs) for different types of task interaction (VE vs. RE). However, there was significant correlation of AMAs measured in the RE vs. VE, indicating that despite the reported discrepancies between the VE and RE, the trends of the AMAs were the same across the various task manipulations. This conclusion is consistent with the findings of the present research, which provided physiological evidence supporting use of the VR-BD task in a manner similar to the native task.

The third hypothesis (H3) stated that greater muscle activation levels would be generated during performing rotary movement versus simple transitional movements of blocks. This expectation was supported by the experimental results. All four muscles showed significantly larger responses during the rotation phase as compared to the honing phase. In the study by Crossman and Goodeve (1983), it was shown that an index of performance in a

1-DOF rotary task was comparable with that in translational movements. Stoelen et al. (2010) also confirmed that combined movements with rotations and translations of 1-DOF can be approximated using a Fitts' Law-equivalent of the task. However, in modern human-machine interface applications, users are required to control up to 3-DOF of rotational movement and 3-DOF of translational movement, simultaneously. Examples include the manipulation of virtual objects in computer-aided modeling, virtual reality applications, and teleoperation of robot manipulators in which an operator typically controls the translation and rotation of a robot end effector. The VR-simulation of the BD task in the present study is another example. Considering the fact that the rotary task involved in BD test required 3-DOF movement, the difference in results among studies suggests the original form of Fitts' Law is not applicable for evaluating performance involving more complicated rotations. In research on multidimensional formulations of Fitts' Law, it was found that the index of performance in rotational movements was similar to that found in translational movements (Crossman & Goodeve, 1983) and expanded formulations could represent rotational tasks reasonably well (Stoelen et al., 2010).

The fourth hypothesis (H4) posited that different activation levels would occur among the four muscles monitored in the experiment. This contention was partially supported by the experimental results. Comparisons among muscle responses revealed the flexor pollicis brevis to produce significantly higher activation levels as compared to the other three muscles under both test conditions. This result was due to the important role the thumb played during holding and rotating blocks (or the block-shaped interface at the haptic device).

Beyond this finding, there was an interesting trend in muscle activation between the test conditions. On average, the pronator teres, flexor carpi ulnaris and extensor carpi radilis all showed higher responses under the VR condition. However, opposite to this, the flexor pollicis brevis showed relatively higher activation under the native test condition. Such difference in the pattern of performance suggested that the flexor pollicis brevis might be used more often in performing the native BD task, which was likely a result of physical task behaviors. Participants could easily control blocks with the thumb and slight movements of other fingers, especially during the rotation phase. While doing this, it was possible for participants to hold their elbows still. However, in order to manipulate a virtual block with the haptic control device, participants usually had to drag the block-shaped control in order to position the virtual block at the target grid area. In addition, while rotating virtual blocks with the haptic control, the required magnitude of movement of the forearm was larger than in the native task. Therefore, both the thumb and forearm were used to move the virtual block to the target area with the desired orientation. In addition, the haptic device had a rotational limitation preventing participants from rotating a block in only one direction in order to achieve all possible faces for placement. This behavior was, however, possible with a physical block in hand. With such a constraint, participants had to twist their forearms while rotating the block-shaped control device in order to complete the VR-based task.

Considering the fact that the set of muscles observed in this study revealed no substantial differences in activation between the VR and native conditions, it is possible that in the VR test condition, the three forearm muscles served as compensation to the loss of freedom at thumb. This situation likely alleviated otherwise heavy reliance on the flexor pollicis brevis, observed in native test performance. This finding is important in understanding the potential causes of different performance under the native and VR test conditions. It also provides us with some information on how to further improve the VR-BD task as a simulation of native task performance.

The fifth hypothesis (H5) posited that comparable NASA-TLX scores would be observed for BD task performance under the native and VR test conditions. This hypothesis was partially supported by the experimental results. Contrary to the expectation, the overall TLX scores were higher for participants under the VR-condition, indicating that they perceived higher workload while performing the computerized BD task. However, analysis of ratings for each dimension of the NASA-TLX provided more interesting information. It was found that there was no significant difference in perceived mental demand among the test conditions, indicating that the VR-BD simulation did not require additional mental and perceptual activity, such as thinking, deciding and remembering. The VR-BD simulation did not make the tasks seem more demanding or complex to the participants. This result was consistent with hypothesis on the basis of prior performance results in use of the VR-BD simulation and native task; however, it was inconsistent with the study of Pontonnier et al. (2014). In general, they found task performance in their VE to be more difficult. In their study, task difficulty was defined as the ease with which participants recognized assembly pieces and the ease of performance of the assembly task. At the end of the experiment, they asked participants to rate on a five-point scale the level of difficulty of the tasks in both RE and VE. Therefore, the perceived higher difficulty of tasks in the VE may be confounded by other factors, such as participant frustration, and should not be considered as just more complex mental processing. However, no force feedback was provided in the Pontonnier et al. system, which was identified as the main reason for higher perceived task difficulty. Therefore, reliable force feedback features of the haptic-control in the present study probably made a contribution to the comparability of the VR and native task conditions.

No significant difference in performance ratings were found among the test conditions, indicating that participants assigned to the VR condition were as satisfied with their performance as those assigned to the native condition.

It is also important to note here that no significant difference was shown in physical demand ratings. This suggested that the VR-BD task did not seem more laborious to participants (despite any differences in time-to-task completion). The VR-BD simulation did not require additional physical effort in controlling the virtual blocks using the block-shaped haptic device, supporting equivalence of the computerized version of the task with the native BD task in terms of physical demand.

With respect to the significantly higher overall TLX scores for participants in the VR group, temporal demand ratings were also higher, indicating that participants perceived greater time pressure in the virtual test. This finding was consistent with the results on task completion time; participants under the VR condition spent more time in completing the BD task. Related to this, higher ratings of effort in the VR-BD task indicated that the simulation seemed more challenging to participants. Frustration ratings were also higher for participants under the VR conditions, likely to the above issue with poor block placement at the target grid. Such ratings indicate that VR participants felt more insecure, discouraged, irritated or stressed while interacting the simulated task. The manner in which participants were required to control blocks in the VR simulation was not as straightforward as controlling physical blocks with the hand. In addition, as previously mentioned, the perception of distance in the "3D" simulation was more challenging than in physical reality. These reasons all could have possibly caused interaction with the VR simulation to differ from participant expectations, leading to frustration. Pontonnier et al. (2014) assumed that cognition was affected by the types of environment (VE vs. RE) and motor control was also altered. Altered cognition induces altered motor control in task performance in an unfamiliar environment. For the present research, it is possible that participant frustration could be eliminated if greater simulation familiarization time or practice trials were provided.

Strong correlations were found among some of experiment response measures. Under the native test condition, the overall TLX score and some of the sub-dimension ratings were found to be correlated with the flexor carpi ulnaris (CH3) activation level; whereas, under the VR condition, correlations of the workload measures were found with pronator teres (CH2) activation levels. This result indicated that the pronator teres played a more important role in interaction with the VR-BD simulation than it did in the physical BD task performance.

Under the native test condition, task time was found to be positively correlated with pollicis brevis (CH1) and extensor carpi radialis (CH4) activation levels, meaning that the TTC was longer when the participant held the block and moved the block with more muscle effort. However, under VR test condition, TTC was not correlated with pollicis brevis (CH1) activation level in the rotation phase of movement, indicating that the CH1 muscle played a less important role in rotating the virtual blocks. These results provide additional evidence that participants used the forearm muscles in the VR test condition to compensate for thumb muscle use while rotating blocks.

Under the native test condition, mental demand and frustration scores were positively correlated with TTC for all trials, indicating that more time-consuming tasks posed higher mental demand and made participants more discouraged or stressed. Under the VR condition, TTC was positively correlated with performance ratings as a sub-dimension of TLX scores, indicating that participants mainly used task completion time as a basis to evaluate their performance.

Although muscle activation level and ratings of some sub-dimensions of the NASA-TLX appeared to be comparable under the two test conditions, the different patterns of response correlations between the native and VR setups indicated that there might still be some differences in upper-extremity muscle use among the conditions and, consequently, participants' perceptions of workload in the tests. In addition, although EMG responses of forearm muscles (i.e., CH2, CH3 and CH4) are significantly lower than that of thumb muscle (CH1), the correlation results showed that subjects perceived more workload when using forearm muscles during VR conditions. These results also provided additional evidence that subjects use forearm muscles to compensate the restriction of the use of thumb muscle under the VR condition.

8. Conclusion

The objective of this research was to assess the validity of a VR-based simulation of a standardized psychomotor test from a biometric perspective. In contrast to prior investigations focusing on performance based validations, this study focused on muscle responses as well as user perceived cognitive workload as complementary measures of simulation evaluation.

A between-subjects experiment was conducted in which performance under two test conditions (VR vs. native) was compared on the basis of the muscle activation (EMG) responses of four muscles used in task performance, participant task completion time, and subjective TLX workload ratings. Results of the experiment suggested that levels of forearm muscle activation were comparable across the two test conditions (native BD task vs. VR-BD task). However, task completion time with VR-BD simulation was still higher than with the native BD task, likely due to fine differences in physical interaction styles with the two versions of the task and participant difficulty in perceiving distances in the VR simulation. Correlation analyses provided some evidence in support of these explanations. Results of the study also revealed comparable participant perceptions of mental demand, physical demand and performance ratings as part of the NASA-TLX; however, different levels of temporal demand, participant frustration and effort ratings as well as overall TLX scores were also observed in the study.

In general, the study showed comparable process-oriented measurements (i.e., EMG responses) among the two test task conditions along with some similarities in product-oriented measures (i.e., specific demand components of the NASA-TLX). Having said this, there were also some differences in terms of gross product-oriented measures, including overall task completion time and workload score.

Participant performance was considered to be consistent across test conditions, although they claimed higher workload with the VR-BD simulation. Therefore, it can be concluded from this study that it is possible to develop computerized-versions of standardized psychomotor tests to facilitate objective and systematic evaluation of motor control as well as to provide a platform for training and learning of fine motor skills from a process perspective. Another conclusion is that custom design of haptic control interfaces can lead to levels of muscle responses in VR-based simulation of psychomotor tasks comparable to native task performance. Therefore, the computer-based simulation evaluated in the present study may allow users to achieve similar levels of skill rehabilitation as native tasks.

8.1 Applications

The VR-BD simulation system represents a major advance over current clinical practice of subjective evaluation of patient motor control capability. More accurate and objective data can be obtained through the computer software. A broad set of kinematic responses can be captured, including haptic control displacement, velocity and acceleration as well as forces at the interface, which was not possible with traditional task administration. A great amount of time and human labor could be saved if such computerized tests could be used as substitutes to physical versions. Consequently, the computerized BD task may play an important in psychomotor skill rehabilitation.

As an example of the use of VEs for delivery of rehabilitation tasks, Patton et al. (2001) designed a study that involved hemiparetic subject (persons with stroke) control of a planar manipulandum with a handle that was subjected to a virtual force field. They found that hemiparetic individuals could learn movement patterns lost to stroke. Thus, they demonstrated the potential for the novel use of a VE to teach motor skills lost after injury to the brain. Furthermore, given the ease of use of computerized tests, similar to the VR-BD simulation used in the present study, skill rehabilitation processes can be precisely controlled and tracked.

Beyond these applications, previous research (Aggarwal et al., 2007) has shown that proficiency in a VR-based training curriculum shortened the learning curve for abled users in a real laparoscopic procedure when compared with traditional training methods. Similarly, another application of the simulation in the present study could be motor skill development in abled persons for complex assembly tasks in industrial environments. Ma (2014) recently found the VR-BD simulation to have utility for worker training in manual assembly requiring fine handwork with small objects (e.g., iPhone assembly). Furthermore, it is possible to classify individual motor ability level with data from such computerized psychomotor performance tests. This information can be used as a basis for prescribing motor control training regimens for specific levels of performance in assembly tasks (Ma, 2014).

8.2 Limitations

As identified in the Discussion section, the VR system applied in the current study had limitations in terms of simulating physical hand control of objects as well as supporting perceptions of physical distances in virtual display space. First, participants viewed the VR interface through a computer screen with 3D glasses. Thus, perceiving distance and location of the virtual blocks occurred without accommodation and resulted in greater time to perform the task. Second, the haptic device had a rotational limitation preventing participants from rotating a block in one direction in order to achieve all possible faces. This type of rotation was possible with a physical block in hand. Third, the handhold control for virtual blocks was attached to the Omni haptic device. Although the control was customized in design to the same shape and size of physical blocks, slight resistance and lack of smoothness in movement could be perceived while manipulating virtual blocks in the simulation. As Pontonnier et al. (2014) stated, using VR for ergonomic evaluation of assembly tasks is still challenging, as interaction and simulation choices deeply affect sensory feedback, cognition, and motor control. Therefore, VR simulation with higher fidelity and realism is necessary for further computerized psychomotor test development and research. However, many current limitations and drawbacks of VR are associated with hardware technology and are expected to be resolved within the next few years (Rebelo, 2012).

Another limitation of the present study is that only 40 trials of dice-manipulation training were provided to familiarize participants with the VR system. Although this training method previously proved efficient in familiarizing participants with the haptic device (Clamann et al., 2013), it was possible that participants still experienced difficulty and frustration with the control interface during the present experiment. Participants might have needed more training to become mentally accustomed to the interaction style and develop high proficiency in use of the block-shaped control to manipulate virtual blocks.

8.3 Future work

On the basis of the above research limitations, any future work towards validating computerized versions of psychomotor tests should ensure highly realistic simulations in order to achieve comparable, if not superior, product-oriented outcomes as original test designs. It would also be interesting to explore the VR-BD simulation with a different control device that could better simulate the native version of the BD task, e.g., a data glove providing multi-finger feedback in virtual block handling.

Beyond this, further investigation should be conducted on how visual aids and haptic force-feedback can be effectively introduced into the VR simulation in order to direct user attention and motor behavior. It is possible that such aids may serve to further accelerate motor task training and learning under specific therapy regimens.

Finally, the present research focused on the well-known BD task as part of the adult IQ test. The BD test is one of the most prevalent psychomotor tests in clinical use. It would be beneficial if computerized versions of other tests, such as the Matrix Reasoning component of the adult IQ test, could be developed and validated with various reliable measures, including performance, muscle activation levels and cognitive workload responses.

REFERENCES

Aggarwal, R., Ward, J., Balasundaram, I., Sains, P., Athanasiou, T., & Darzi, A. (2007). Proving the effectiveness of virtual reality simulation for training in laparoscopic surgery. *Annals of surgery*, *246*(5), 771-779.

Backs, R. W., & Walrath, L. C. (1992). Eye movement and pupillary response indices of mental workload during visual search of symbolic displays. *Applied ergonomics*, *23*(4), 243-254.

Benedetto, S., Pedrotti, M., Minin, L., Baccino, T., Re, A., & Montanari, R. (2011). Driver workload and eye blink duration. *Transportation research part F: traffic psychology and behaviour*, *14*(3), 199-208.

Berchou, R., & Block, R. I. (1983). Use of computerized psychomotor testing in determining CNS effects of drugs. *Perceptual and motor skills*,*57*(3), 691-700.

Brown, I. D. (1978). Dual task methods of assessing work-load. *Ergonomics*,*21*(3), 221-224.

Casellato, C., Pedrocchi, A., Zorzi, G., Vernisse, L., Ferrigno, G., & Nardocci, N. (2013). EMG-based visual-haptic biofeedback: a tool to improve motor control in children with primary dystonia. *Neural Systems and Rehabilitation Engineering, IEEE Transactions on*, *21*(3), 474-480.

Causby, R., Reed, L., Mcdonnell, M., & Hillier, S. (2014). Use of Objective Psychomotor Tests in Health Professionals. *Perceptual & Motor Skills*, *118*(3), 765-804.

Clamann, M., & Kaber, D. B. (2012, September). The effects of haptic and visual aiding on psychomotor task strategy development during virtual reality-based training. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 56, No. 1, pp. 2570-2574). SAGE Publications.

Clamann, M., Ma, W., & Kaber, D. B. (2013, October). Evaluation of a Virtual Reality and Haptic Simulation of a Block Design Test. In *Proceedings of the 2013 IEEE International Conference on Systems, Man, and Cybernetics* (pp. 882-887). IEEE Computer Society.

Clamann, M., Ma, W., & Kaber, D. (2014). Comparison of haptic control design for virtual reality-based assembly task training. IIE Annual Conference.Proceedings, , 1391-1400.Comrey, A. L. (1953). Group performance in a manual dexterity task. *Journal of Applied Psychology*, *37*(3), 207.

Costa, L. D., Vaughan Jr, H. G., Levita, E., & Farber, N. (1963). Purdue Pegboard as a predictor of the presence and laterality of cerebral lesions.*Journal of consulting psychology*, *27*(2), 133.

Crossman, E. R. F. W., & Goodeve, P. J. (1983). Feedback control of hand-movement and Fitts' law. *The Quarterly Journal of Experimental Psychology*,*35*(2), 251-278.

Diller, L., Weinberg, J., Gordon, W., Goodkin, R., Gerstman, L. J., & Ben-Yishay, Y. (1974). *Studies in cognition and rehabilitation in hemiplegia*.

Engelhorn, R. (1988). EMG and motor performance changes with practice of a forearm movement by children. *Perceptual and motor skills*, *67*(2), 523-529.

Fillmore, M. T. (2003). Reliability of a computerized assessment of psychomotor performance and its sensitivity to alcohol-induced impairment.*Perceptual and motor skills*, *97*(1), 21-34.

Fillmore, M. T., & Vogel-Sprott, M. (1994). Psychomotor performance under alcohol and under caffeine: Expectancy and pharmacological effects.*Experimental and Clinical Psychopharmacology*, *2*(4), 319.

Fillmore, M. T., & Vogel-Sprott, M. (1995). Expectancies about alcohol-induced motor impairment predict individual differences in responses to alcohol and placebo. *Journal of Studies on Alcohol and Drugs*, *56*(1), 90.

Fillmore, M. T., & Vogel-Sprott, M. (1998). Behavioral impairment under alcohol: cognitive and pharmacokinetic factors. *Alcoholism: Clinical and Experimental Research*, *22*(7), 1476-1482.

Florimond, V. (2009). Basics of surface electromyography applied to physical rehabilitation and biomechanics. *Montreal, Canada: Thought Technology Ltd*.

Gallagher, C., & Burke, T. (2007). Age, gender and IQ effects on the Rey-Osterrieth Complex Figure Test. *British Journal of Clinical Psychology*, *46*(1), 35-45.

Gath, I., & Stalberg, E. V. (1976). Techniques for improving the selectivity of electromyographic recordings. *Biomedical Engineering, IEEE Transactions on*, (6), 467-472.

Göbel, M. (2005). Electromyography (EMG). In Stanton, N. A., Hedge, A., Brookhuis, K., Salas, E., & Hendrick, H. W. (Eds.), *Handbook of human factors and ergonomics methods (pp. 19-1 to 19-8)*. CRC Press.

Goldstein, K., & Scheerer, M. (1941). Abstract and concrete behavior an experimental study with special tests. *Psychological monographs*, *53*(2), i.

Grant, R. C., Carswell, C. M., Lio, C. H., & Seales, W. B. (2013). Measuring surgeons' mental workload with a time-based secondary task. *Ergonomics in Design: The Quarterly of Human Factors Applications*, *21*(1), 7-11.

Grote, C., & Salmon, P. (1986). Spatial complexity and hand usage on the block design test. *Perceptual and motor skills*, *62*(1), 59-67.

Hägg, G. M., & Milerad, E. (1997). Forearm extensor and flexor muscle exertion during simulated gripping work—an electromyographic study. *Clinical Biomechanics*, *12*(1), 39-43.

Hart, S., & Staveland, L. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. Human Mental Workload (pp. 139-183). Amsterdam: North Holland.

Helander, M. (1978). Applicability of drivers' electrodermal response to the design of the traffic environment. *Journal of applied Psychology*, *63*(4), 481.

Hertzog, C., Lineweaver, T. T., & Hines, J. C. (2014). COMPUTERIZED ASSESSMENT OF AGE DIFFERENCES IN MEMORY BELIEFS 1, 2.*Perceptual & Motor Skills*, *119*(2), 609-628.

Hill, S. G., Iavecchia, H. P., Byers, J. C., Bittner, A. C., Zaklad, A. L., & Christ, R. E. (1992). Comparison of four subjective workload rating scales. Human Factors, 34(4), 429-439.

HOCKEY, G. R. J., Briner, R. B., Tattersall, A. J., & Wiethoff, M. (1989). Assessing the impact of computer workload on operator stress: the role of system controllability. *Ergonomics*, *32*(11), 1401-1418.

Hu, B., Ma, L., Zhang, W., Salvendy, G., Chablat, D., & Bennis, F. (2011). Predicting real-world ergonomic measurements by simulation in a virtual environment. *International Journal of Industrial Ergonomics*, *41*(1), 64-71.

Jeon, W., Clamann, M., Kaber, D. B., & Currie, N. J. (2013, October). Assessing Goal-Directed Three-Dimensional Movements in a Virtual Reality Block Design Task. In *Systems, Man, and Cybernetics (SMC), 2013 IEEE International Conference on* (pp. 3739-3744). IEEE.

Jorna, P. G. A. M. (1992). Spectral analysis of heart rate and psychological state: A review of its validity as a workload index. *Biological psychology*,*34*(2), 237-257.

Judge, J., & Stirling, J. (2003). Fine motor skill performance in left-and right-handers: Evidence of an advantage for left-handers. *Laterality*, *8*(4), 297-306.

Kaber, D., Tupler, L. A., Clamann, M., Gil, G. H., Zhu, B., Swangnetr, M., ... & Lee, Y. S. (2014). Evaluation of an augmented virtual reality and haptic control interface for psychomotor training. *Assistive Technology*, *26*(1), 51-60.

Kane, J., & Gill, R. P. (1972). Implications of the Purdue pegboard as a screening device. *Journal of Learning Disabilities*, *5*(1), 36-40.

Kohs, S. C. (1923). Intelligence measurement: A psychological and statistical study based upon the block-design tests.

Kramer, A. F., Trejo, L. J., & Humphrey, D. G. (1996). Psychophysiological measures of workload- Potential applications to adaptively automated systems.*Automation and human performance: Theory and applications(A 98-12010 01-54), Mahwah, NJ, Lawrence Erlbaum Associates, Publishers, 1996,*, 137-162.

Li, Y., Kaber, D. B., Tupler, L., & Lee, Y. S. (2011). Haptic-based Virtual Environment Design and Modeling of Motor Skill Assessment for Brain Injury PatientsRehabilitation. *Computer-Aided Design and Applications*, *8*(2), 149-162.

Li, Y., Clamann, M., & Kaber, D. B. (2013). Validation of a haptic-based simulation to test complex figure reproduction capability. *Human-Machine Systems, IEEE Transactions on*, *43*(6), 547-557.

Lin, C. K., & Wu, H. M. (2014). Development and validation of the computerized bilateral motor coordination test. *Research in developmental disabilities*, *35*(1), 110-116.

Luzzi, S., Pesallaccia, M., Fabi, K., Muti, M., Viticchi, G., Provinciali, L., & Piccirilli, M. (2011). Non-verbal memory measured by Rey–Osterrieth Complex Figure B: normative data. *Neurological Sciences*, *32*(6), 1081-1089.

Ma, W. (2014). An Application of Quantitative Methods for Motor Ability Level Classification, Performance Prediction and Training Protocol Selection.

Marras, W. S. (1990). Industrial electromyography (EMG). *International Journal of Industrial Ergonomics*, *6*(1), 89-93.

Nygren, T. E. (1991). Psychometric properties of subjective workload measurement techniques: Implications for their use in the assessment of perceived mental workload. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *33*(1), 17-33.

O'Donnell, R.D. & Eggemeier, F.T. (1986). "Workload assessment methodology". In K.R. Boff, L. Kaufman & J.P. Thomas (Eds.), Handbook of perception and human performance. Volume II, cognitive processes and performance. (pp 42/1-42/49). New York: Wiley

Ogino, T., Watanabe, K., Nakano, K., Kado, Y., Morooka, T., Takeuchi, A., ... & Ohtuska, Y. (2009). Predicting executive function task scores with the Rey-Osterrieth Complex Figure. *Brain and Development*, *31*(1), 52-57.

Patton, J. L., Mussa-Ivaldi, F. A., & Rymer, W. Z. (2001). Altering movement patterns in healthy and brain-injured subjects via custom designed robotic forces. In *Engineering in Medicine and Biology Society, 2001. Proceedings of the 23rd Annual International Conference of the IEEE* (Vol. 2, pp. 1356-1359). IEEE.

Pontonnier, C., Samani, A., Badawi, M., Madeleine, P., & Dumont, G. (2014). Assessing the Ability of a VR-Based Assembly Task Simulation to Evaluate PhysicalRisk Factors. *Visualization and Computer Graphics, IEEE Transactions on*, *20*(5), 664-674.

Rebelo, F., Noriega, P., Duarte, E., & Soares, M. (2012). Using virtual reality to assess user experience. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 0018720812465006.

Reid, G. B., & Nygren, T. E. (1988). The subjective workload assessment technique: A scaling procedure for measuring mental workload. *Advances in psychology*, *52*, 185-218.

Reiner, M., & Gelfeld, T. M. (2014). Estimating mental workload through event-related fluctuations of pupil area during a task in a virtual world. *International Journal of Psychophysiology*, *93*(1), 38-44.

Rocha-Amador, D., Navarro, M., Trejo-Acevedo, A., Carrizales, L., Pérez-Maldonado, I., Díaz-Barriga, F., & Calderón, J. (2009). Use of the Rey-Osterrieth Complex Figure Test for neurotoxicity evaluation of mixtures in children. *Neurotoxicology*, *30*(6), 1149-1154.

Roscoe, A. H. (1992). Assessing pilot workload. Why measure heart rate, HRV and respiration?. *Biological psychology*, *34*(2), 259-287.

Rozencwajg, P., & Corroyer, D. (2002). Strategy development in a block design task. *Intelligence*, *30*(1), 1-25.

Sabourin, M., & Rioux, S. (1979). Effects of active and passive EMG biofeedback training on performance of motor and cognitive tasks. *Perceptual and motor skills*, *49*(3), 831-835.

Schorr, D., Bower, G. H., & Kiernan, R. J. (1982). Stimulus variables in the block design task. *Journal of consulting and clinical psychology*, *50*(4), 479.

Sahakian, B. J., & Owen, A. M. (1992). Computerized assessment in neuropsychiatry using CANTAB: discussion paper. *Journal of the Royal Society of Medicine*, *85*(7), 399.

Salthouse, T. A. (1987). Sources of age-related individual differences in block design tests. *Intelligence*, *11*(3), 245-262.

Sartorio, F., Bravini, E., Vercelli, S., Ferriero, G., Plebani, G., Foti, C., & Franchignoni, F. (2013). The Functional Dexterity Test: Test–retest reliability analysis and up-to date reference norms. *Journal of Hand Therapy*, *26*(1), 62-68.

Schorr, D., Bower, G. H., & Kiernan, R. J. (1982). Stimulus variables in the block design task. *Journal of consulting and clinical psychology*, *50*(4), 479.

Shin, M. S., Park, S. Y., Park, S. R., Seol, S. H., & Kwon, J. S. (2006). Clinical and empirical applications of the Rey–Osterrieth complex figure test.*Nature protocols*, *1*(2), 892-899.

SIREVAAG, E. J., KRAMER, A. F., REISWEBER, C. D. W. M., STRAYER, D. L., & GRENELL, J. F. (1993). Assessment of pilot performance and mental workload in rotary wing aircraft. *Ergonomics*, *36*(9), 1121-1140.

Song, J., Zhuang, D., Song, G., & Wanyan, X. (2011, October). Pilot mental workload measurement and evaluation under dual task. In *Biomedical Engineering and Informatics (BMEI), 2011 4th International Conference on*(Vol. 2, pp. 809-812). IEEE.

Sousa, A. S., & Tavares, J. M. R. (2012). Surface electromyographic amplitude normalization methods: a review. *Electromyography: New Developments, Procedures and Applications*.

Spreen, O., & Benton, A. L. (1965). Comparative studies of some psychological tests for cerebral damage. *The Journal of nervous and mental disease*, *140*(5), 323-333.

Stefanidis, D., Korndorffer, J. R., Black, F. W., Dunne, J. B., Sierra, R., Touchard, C. L., ... & Scott, D. J. (2006). Psychomotor testing predicts rate of skill acquisition for proficiency-based laparoscopic skills training. *Surgery*,*140*(2), 252-262.

Stoelen, M. F., & Akin, D. L. (2010). Assessment of Fitts' Law for quantifying combined rotational and translational movements. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *52*(1), 63-77.

Telles, S., Yadav, A., Kumar, N., Sharma, S., Visweshwaraiah, N. K., & Balkrishna, A. (2013). Blood pressure and purdue pegboard scores in individuals with hypertension after alternate nostril breathing, breath awareness, and no intervention. *Medical Science Review*, *19*, 61-66.

Tiffin, J., & Asher, E. J. (1948). The Purdue Pegboard: norms and studies of reliability and validity. *Journal of applied psychology*, *32*(3), 234.

Towles, J. D. (2004, March). Performing cadaveric experiments on the thumb to determine the potential contribution of thumb muscles to the endpoint force during grasping. In *Haptic Interfaces for Virtual Environment and Teleoperator Systems, 2004. HAPTICS'04. Proceedings. 12th International Symposium on*(pp. 347-350). IEEE.

Wechsler, D. (1958). The Measurement and Appraisal of Adult Intelligence .

Wechsler, D. (1999). Wechsler Abbreviated Scale of Intelligence. Psychological Corporation, San Antonio, TX.

Whitman, L. E., Jorgensen, M., Hathiyari, K., & Malzahn, D. (2004, December). Virtual reality: its usefulness for ergonomic analysis. In*Proceedings of the 36th conference on Winter simulation* (pp. 1740-1745). Winter Simulation Conference.

Yiu, E. M., Verdolini, K., & Chow, L. P. (2005). Electromyographic study of motor learning for a voice production task. *Journal of Speech, Language, and Hearing Research*, *48*(6), 1254-1268.

Young, M. S., & Stanton, N. A. (2005).Mental Workload. In Stanton, N. A., Hedge, A., Brookhuis, K., Salas, E., & Hendrick, H. W. (Eds.), *Handbook of human factors and ergonomics methods (pp. 39-1 to 199)*. CRC Press.

Zhang, J. H., & Wang, X. Y. (2008). Use of heart rate variability analysis for quantitatively assessing operator's mental workload. In *2008 International Conference on BioMedical Engineering and Informatics* (Vol. 1, pp. 668-672).

APPENDICES

Appendix A Informed Consent Form

North Carolina State University
INFORMED CONSENT FORM for RESEARCH

Title of Study: BIOMECHANICAL COMPARISON OF BLOCK DESIGN TO VIRTUAL REALITY

Principal Investigator: Michael Clamann, Linus Jeon, Wenqi Ma, May Swangnetr, Wenjuan Zhang, Maicom Brandao
Faculty Sponsor (if applicable): David Kaber

General Information
You are being asked to take part in a research study. Your participation in this study is voluntary. You have the right to be a part of this study, to choose not to participate or to stop participating at any time. The purpose of this research study is to gain a better understanding of how controls are used in virtual reality-based (VR) simulations of motor tasks for motor control skill development. You are not guaranteed any personal benefits from being in this study. Research studies like this may also pose risks to those that participate. In this consent form you will find specific details about the research in which you are being asked to participate. If you do not understand something in this form it is your right to ask for clarification or more information. A copy of this consent form will be provided to you.

Purpose of this study
The purpose of this study is to learn more about how to develop effective controls for interaction with VR systems for motor skill training. The VR simulation to be used in this study involves a visual display on a desktop monitor with a 3-dimensional perspective. Your control of events in the VR simulation will occur through a hand-held custom "block" control (similar to a joystick; see below image) that provides force-feedback (when a virtual cursor contacts a virtual object). The tasks to be performed in this study involve simulated shape manipulation and assembly. Your performance while using the VR system is expected to provide insights into VR design features.

Procedure
If you agree to participate in this study, you will first be asked to read and sign this consent form. You should ensure that you understand all aspects of the research before agreeing. Subsequently, you will be asked to

complete an experimental session using the VR system or a physical replica of the simulated task. This session will last approximately 3.5 hours and will involve training and testing in the native and VR versions of the task through several trials. Four electromyography (EMG) sensors will be attached to your forearm using rubbing alcohol (for cleaning the skin surface), hypo-allergenic electrode gel, adhesive washers, and medical tape. Conductive gel will be added to the electrodes to improve communication with the EMG data collection system. Upon completion of test trials, you will be asked to assess workload using a standardized index developed by NASA. All training and test trials will be videotaped for experimenter data collection purposes but your face will not appear in any video images.

Risks
Risks from this research include: (1) general fatigue due to attending to the VR displays (or block patterns) during the test trials; (2) slight discomfort due to wearing the EMG sensors; (3) potential skin irritation upon removal of EMG electrodes. To address these risks, you will be provided with rest periods between trials, experimenters will attend to any skin reaction due to alcohol exposure, and your skin may be shaved to remove hair in advance of electrode placement.) If you have sensitive skin or are allergic to alcohol, please inform the researchers now. If you do not have such sensitivities or allergies, please mark your initials here: _____.

Benefits
There are no direct benefits to you from this research. You may gain some knowledge about how contemporary VR systems are being used for diagnostic testing and training of motor skills for persons suffering from minor traumatic brain injuries, etc.

Eligibility
You must be 18 years or older on the date of your experiment session to participate in this study. You must also be right-handed, defined by a score above 90 on the Edinburgh Handedness Inventory, and have no major upper extremity impairments affecting either hand. You will need to complete the handedness inventory in advance of task training and testing. Please initial here to confirm that you have no major upper extremity impairment: _____.

Confidentiality
The information in the study records will be kept confidential. Data will be stored securely in the Cognitive Ergonomics Lab in the NC State Department of Industrial and Systems Engineering. The data will only be made available to the persons conducting the study. No reference will be made to you in oral or written reports of the study, which could link you to the research. A background survey will collect identifying information such as your name. Other data will include your gender, age, etc. and will be used for demographic statistics. A code number will be written on your background survey and matched to your name. The background surveys for all participants will be kept separately from all other response data collected as part of the experiment. The surveys and all other data will be destroyed at the close of the study. All video recordings to be used for data analysis will also be destroyed at the conclusion of the study.

Compensation
For participating in this study you will be paid $15 per hour, for a total of $52.50 (3.5 hours at $15.00 per hour), provided you complete the entire study. If you withdraw from the experiment prior to its completion, you will receive compensation at a rate of $15 per hour for any time that you provided.

Contact
If you have questions at any time about the study or the procedures, you may contact the researcher, Dave Kaber, at 448 Daniels Hall, NC State University main campus, or 919-515-3086. If you feel you have not been treated according to the descriptions in this form, or your rights as a participant in research have been violated during the course of this project, you may contact Deb Paxton, Regulatory Compliance Administrator, Box 7514, NCSU Campus (919/515-4514).

Consent to Participate

"I have read and understand the above information. I have received a copy of this form. I agree to participate in this study with the understanding that I may withdraw at any time."

Participant's signature_______________________________________ **Date** _________________

Investigator's signature__________________________________ **Date** _________________

Appendix B Experiment Procedure

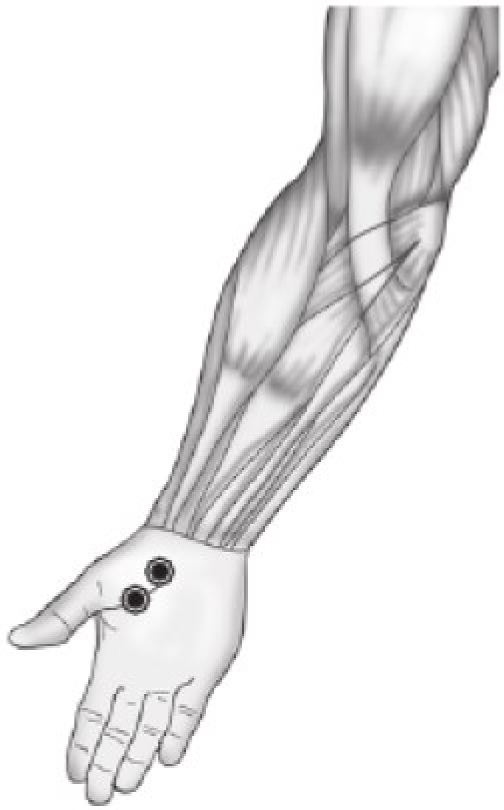
This document presents instructions and scripts to administer the test and therapy sessions for the Electromyography (EMG) Haptic Study. The Instructions are provided in tabular format. Sections to be read to participants appear in *italics*. Supplemental information (e.g., materials, scoring methods) is also provided in each section.

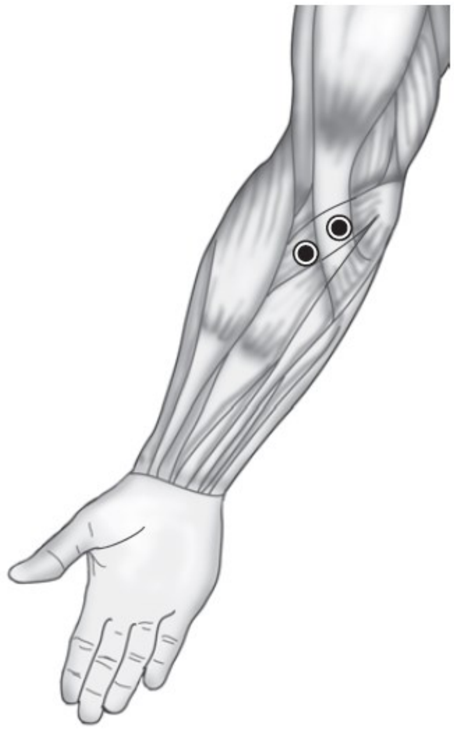
Orientation

Introduction	[Seated the participant in outer side of the room.] *Thank you for participating in this experiment. Today you will perform a psychomotor test, or test that combines cognitive and physical skills. You will use colored blocks to construct a series of design patterns as quickly as possible. The test takes about 30 minutes. The experiment session will last about two and a half hours.* *Please understand this experiment is not to test your personal ability or skills. The goal of this study is to assess the impact of various training tasks on psychomotor performance, independent of ability level.*
Informed consent	[Prepare the informed consent form and a pen.] *This form summarizes information you need to know about the experiment. Please read it. If you have any questions, feel free to ask me. Please note that in order to participate in this study you should be at least 18 years old, have 20/20 vision (with or without correction; glasses or contact lenses are OK) and be right-handed. You will receive $15/hour for your participation. If you consent to participation, please sign and date the form.* [After the participant signs the form, seat them at the test workstation in the outer side of room and begin with the instructions.] *Before we start, please turn off your cell phone. In order to prevent distractions, all electronics need to be off during testing and training. If you need to use the restroom, please do so at this time. You will also be able to take breaks during the experiment, including use of the drinking fountain.*

EMG Prepping	[This part is taken in the outer side of room.] *For the experiment, I will attach 9 electrodes to your forearm and hand to measure muscle activity. The electrodes will be attached to the surface of your skin using adhesive disks similar to a band-aid.* *I will read the procedure for EMG electrode placement before attaching any to your skin.* *First I will determine electrode placement positions on your forearm and mark the locations with a marker. Next, we will shave the locations with an electric shaver. Afterwards, I will clean the locations with rubbing alcohol. I will then apply an abrasive gel to the electrode placement locations to exfoliate the skin. Next, electrodes will be placed at the previously marked spots. Medical tape will then be applied to secure electrodes for testing. Excess electrode cables will be secured to the back of your shirt and seat with tape. After preparing the electrodes, we will commence virtual reality training. During the following experiment, the placement of electrodes may be adjusted to achieve clear signal.* *Do you have any questions?* [Begin the EMG electrode placement procedure. Supplies: • Abrasive gel • EL 504 Electrodes • Marking pen • Pair of scissors • Tape measure • Conductive paste • Shaver • Rubbing alcohol • Medical tape • Cotton ball Steps: • Prepare skin locations for electrode placement. If necessary, shave the area. • Exfoliate the electrode placement site using abrasive gel. • Mark the locations for electrode placement. • Press electrodes onto marked skin location (muscle group and bony

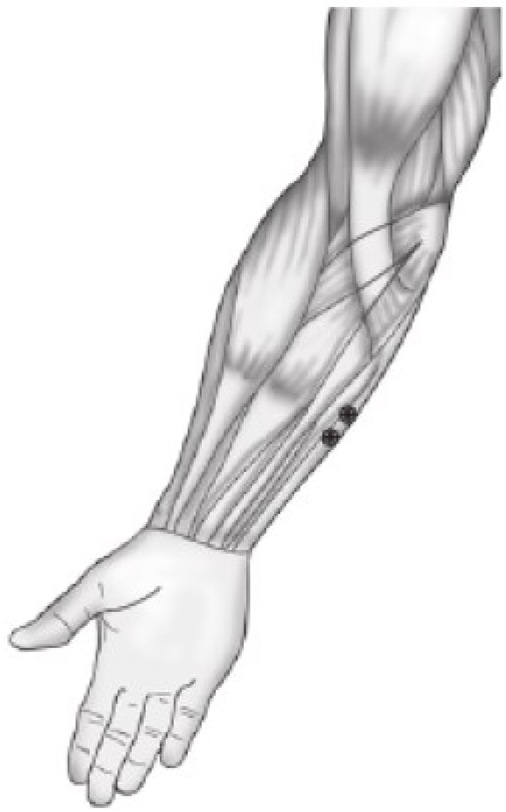
	surface) • Connect leads to the electrodes. • Tape and secure excess cables to the back of participant (make sure the cables will not disturb participant's performance).

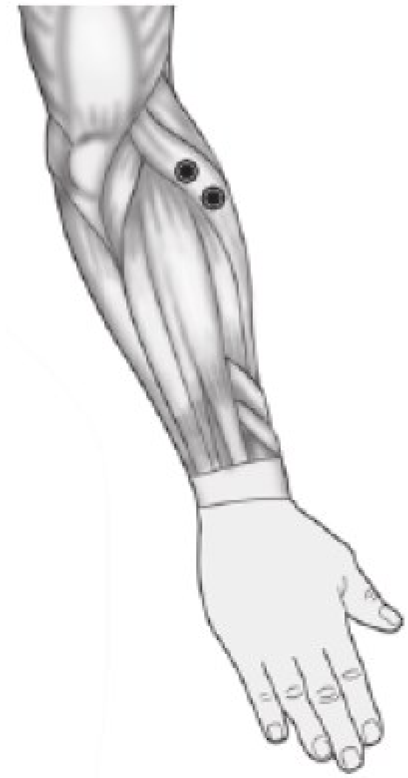
EMG Placement	[EMG signals will be recorded for the following muscle groups: • flexor pollicis brevis, • Pronator teres • extensor carpi radialis and • flexor carpi ulnaris. Precise electrode placement is as follows: **Flexor pollicis brevis** Place electrodes 1 cm apart (cut the electrode adhesive if necessary) on the medial aspect of the thenar eminence, parallel to the direction of the thumb, per the following figure (of the right arm): 

	Pronator teres Support the arm in the palm-up (supinated) position. Palpate in the soft valley in the middle of the ventral aspect of the forearm just below the elbow. Ask the patient to pronate (palm up to palm down) the arm and feel for the muscle mass. Place two active electrodes (2 cm apart) on an oblique angle so that they run parallel to the muscle fibers. See the following figure (of right arm) for an example. 

Flexor carpi ulnaris

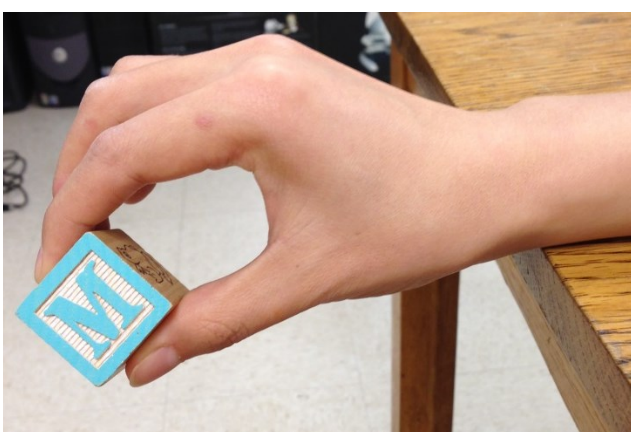
Support the participant's arm while palpating the ventral aspect of the forearm near the elbow on the medial (little finger) side of the arm. Ask the participant to flex the wrist. Place electrodes 2 cm apart over the identified muscle mass in the direction of the muscle fibers. See the following figure (of the right arm) for an example (note: muscle mass occurs towards distal end of forearm):



	Extensor carpi radialis Ask the participant to extend the wrist and locate the muscle mass approximately 5 cm distal from the lateral epicondyle of the elbow, on the dorsal side of the arm just lateral to the brachioradialis. Place two electrodes 2 cm apart over the muscle mass that emerges, with the electrodes running in the direction of the fibers. See the following figure (of the posterior surface of the right arm) for an example: 
	Reference for electrode placement: Criswell, E. (Ed.). (2010). *Cram's introduction to surface electromyography*. Jones & Bartlett Publishers.]

<table>
<tr>
<td>MVC</td>
<td>
[Ask the participant to move to the inner side of the room and seated beside the wooden table. Connect leads to EMG amplifier:

110S-W: white segment to VIN+; black segment to Shield
110S-R: red segment to VIN-; black segment to Shield
110: Ground

Record the baseline maximum voluntary contraction for each muscle group. For each muscle, record activity data and check for clarity of the signal. If the signal is not clear, adjust electrode placement until a clear signal is achieved.
A total of four (4) files should be recorded, one for each muscle group.
Recording MVCs for this study requires a wooden block and a round wooden stick.]
You will now perform Max Voluntary Contractions (MVCs) with your forearm and hand muscles. MVCs are used to establish a baseline for comparison of muscle activity levels under various conditions.
Please sit on the test chair and rest your right forearm on the test table. Please adjust the height of chair and tilt of back rest to make your right upper arm vertical with table surface. During the following MVC part, please keep sitting still and do not move your chair.
[Help the participant to adjust chair position and make markers using erasable marker pen.]
The first test will consist of gripping a wooden block. I will demonstrate.
[Demonstrate the movements as described for the flexor pollicis brevis.]


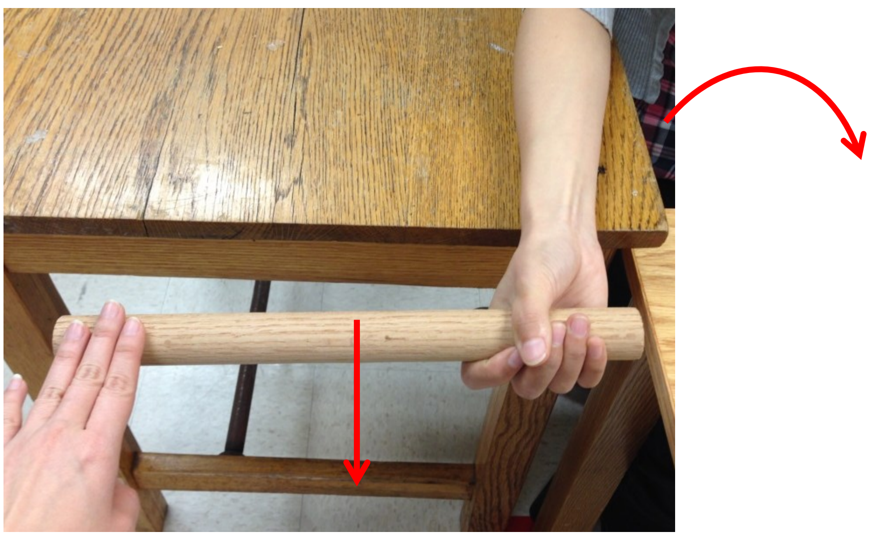
</td>
</tr>
</table>

*Please rest your forearm on the table with you hand and wrist off the edge. With your palm facing downward, grip the block with the index finger and thumb **as tight as possible**. Make sure the pads of the finger and thumb are flat against the centers of the opposite block faces. You will hold this position for 3-5 secs. I will let you know when to release the block.*

Do you have any questions? If not, please squeeze the block like I just did.

I will now perform the second test.

[Demonstrate the movements as described for the pronator teres.]



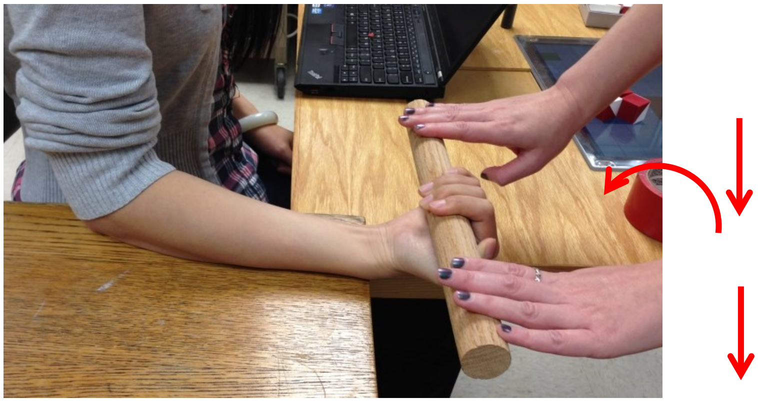
*While sitting in the test chair, rest your forearm on the test table with your hand and wrist off the edge and palm facing up. Hold the round wooden stick in your hand with the longer side of stick on your thumb side. Lift the longer end of stick **as hard as possible** by pronating your forearm. I will apply pressure to the right end of stick as you lift it. Do not lift your arm from the table. Hold the position until I tell you to stop. While doing this movement, please keep your upper arm straight up and down.*

[As the participant lifts the stick, hold down the longer end to resist the motion. Only apply enough pressure to keep the stick parallel to the floor.]

Do you have any questions? If not, please lift the stick.

I will now perform the third test.

[Demonstrate the movements as described for the flexor carpi ulnaris.]



While sitting in the test chair, rest your forearm on the test table with your hand and wrist off the edge and upper arm vertical with table surface. Hold the middle part of the round wooden stick in your hand with your palm facing up. Lift the stick ***as high as possible*** *by flexing your wrist. I will apply pressure to both ends of stick as you lift it. Do not lift your arm from the table. Hold the position until I tell you to stop.*

[As the participant lifts the stick, hold down both ends to resist the motion.]

Do you have any questions? If not, please lift the stick.

Now I will perform the last test.

[Demonstrate the movements as described for the extensor carpi radialis.]

	While sitting in the test chair, rest your forearm on the test table with your hand and wrist off the edge and upper arm vertical with table surface. Hold the middle part of the round wooden stick in your hand with your palm facing down. Lift the stick ***as high as possible*** *by extending your wrist. I will apply pressure to both ends of stick as you lift it. Do not lift your arm from the table. Hold the position until I tell you to stop.* *[As the participant lifts the stick, hold down both ends to resist the motion.]* *Do you have any questions? If not, please lift the stick.*]

Virtual Reality Training

Materials

The virtual reality block design (VR BD) workstation will be used as the training platform. [Ensure that the workstation has the following configuration prior to each session.

1. Monitors set to 800/600 resolution
2. Monitor refresh rate set at 120Hz
3. "\Results" directory is created on the hard drive.]

Procedure

[All participants are to complete VR-based dice manipulation training prior to the VR BD testing. The objective of the dice manipulation training is to provide practice in moving blocks in the VR using a haptic control. The task requires participants to move a single die from the left side of a display screen to a target square on the right side of the screen. The top surface of the die must match a picture of the die located at the top of the screen. Participants should be allowed a 5-minute rest following the dice training.]

Orientation (die task)	*In this phase you will complete a training session to help you become familiar with the equipment you will be using during the experiment. The training session involves 40 trials of a dice manipulation task. The task you'll be performing uses a 3D virtual environment, so you will need to wear active light shutter goggles for all trials.* [Ask the participant to don the 3D glasses. Lift the block-shaped control before starting the trial.] *Now, I am going to ask you to use this block-shaped control to manipulate the die in the virtual environment. Please hold the block as if holding a physical cube. You can use any finger to click the button.* [Show the position of the button on the block.] *Look at the ball-shaped pointer on the virtual table.* [Point to the pointer on the screen.] *Please move the block to control the pointer and position it at the surface of the die. Feel the resistance force when touching the die. You can click the button to hold the die. The button can be released after the die leaves the table and the die will remain attached to the pointer. You can change the die orientation by rotating the block. The die will be automatically released when touching the table. During this process, you are NOT required to keep the same posture in holding the block.* [Demonstrate how to use the block control to manipulate the die.] *The goal of this task is to move the die to the target square shown in black* [point to the target square on the screen] *with the top surface matching the stimulus* [point to the stimulus on the screen]. *Please complete each trial as quickly and accurately as possible.*

	[During training, make sure the AcqKnowledge software is recording EMG signals. The data files will not be analyzed, but this is an opportunity to confirm proper electrode placement.] *Do you have any questions about the training?*
Trial 1 (die task)	*The task you'll be performing uses a 3D virtual environment, so you will need to wear these glasses for all the trials.* [Mark the positions of the chair before starting the program and make sure the chair keeps in the same position before proceeding to next trial in the following session.] [Start the Training program and set the view angle as 0.35 radian (i.e., 20 degrees). Ask the participant to sit straight and keep trunk still while completing the task. Make sure to begin with Trial 1 and confirm the participant is perceiving the VR display as a single image.] *Now you can try to move the die to the target square, matching the top surface to the figure. Try to become familiar with the device as much as possible. We will be recording your time on this task. The timer starts automatically when you pick up the block and stops when you put it down. Keep holding the block control even after the timer stops.* ***You can start when you are ready.***
Trial 2 (die task)	[When the dialog appears at the end of the trial, set the Trial to 2.] *The task is the same. Try to move the die to the square as fast as you can, but be sure the top of the die matches the figure* ***exactly****.* ***You can start when you are ready.***
Trial 3 (die task)	[When the dialog appears at the end of the trial, set the Trial to 3.] *Try to move the die to the square as fast as you can. Be sure the top of the die matches the figure* ***exactly****.* ***You can start when you are ready.***

Trial 4 (die task)	[When the dialog appears at the end of the trial, set the Trial to 4.] *Try to move the die to the square as fast as you can. Be sure the top of the die matches the figure **exactly**.* ***You can start when you are ready.***
	[Following Trial 4, alternate die task training between Trials 3 and 4 until the participant completes 40 trials in total. Allow the participant a 5 minute break when complete.]

Testing

Introduction

[Before starting either the native or the VR BD task, create a new file on the test computer workstation correctly labeled for the task being performed and begin recording data for the experiment.]

Native Materials

[The following equipment and settings are required for each participant:

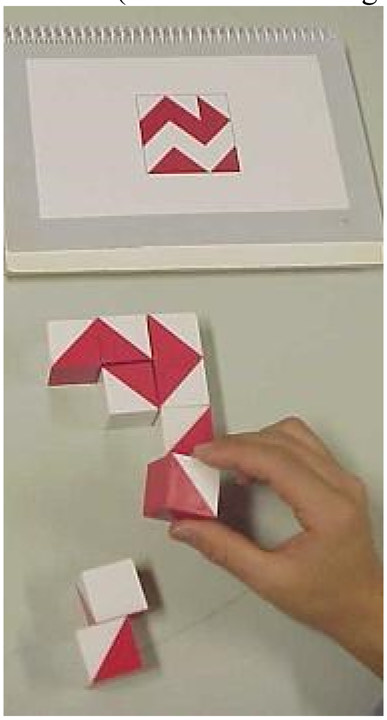
1. A Stopwatch
2. The Wechsler Abbreviated Scale for Intelligence (WASI) stimulus booklet
3. Eight Wechsler blocks.
4. One score sheet identified with a participant number.
5. NASA-Task Load Index (TLX) demand component ranking and rating sheets. Each sheet should be labeled in the same manner as the BD task score sheets
6. The custom VR workstation running with the embedded monitor presenting the "Reality Grid Lines" xls file.
7. The overhead camera focused on the embedded workstation monitor.

Native Procedures

[(Note: All instructions are adapted from the Wechsler Abbreviated Scale for Intelligence (WASI) manual (The Psychological Corporation, 1999).]

The participant must not be permitted to see any of the block designs prior to testing. Designs 1 through 4 are to be presented in two trials. The second trial is administered only if the participant fails to complete the design within the required time period (60 seconds). Design 3 must be demonstrated by the examiner. All other designs (4 through 9) should be presented using the stimulus booklet or by similar means.
Ensure a 7-inch tall workspace between the stimulus booklet and the edge of the table in front of the participant. The booklet or model should be placed slightly to the right of the participant's center (left for left-handed participants). The booklet or model should always be

perpendicular to the participant. The participant is not permitted to move the stimulus booklet. (See the below image for an example of the required layout.)



At the beginning of each trial, blocks should be placed so a variety of surfaces are visible. For the four-block designs, two red sides, one white side and one combination side should be face up. Blocks should never be placed in a single row or column in front of the participant.
Timing starts as soon as you finish reading the instructions for each trial to a participant.
It is acceptable for a participant to revise a design if he or she does so within the time limit (60 sec) and the stopwatch is still running. Credit can only be given for spontaneous corrections.
You may correct the participant once if a design is rotated more than 30 degrees (i.e., "But you see, it goes this way."). Designs rotated after a correction should receive a score of 0.
Participants who fail to complete Designs 3 or 4 on the first trial or fail to complete 3 consecutive designs should be dismissed from the study.
This rule should only be applied for the WAIS-III block designs administered during the test sessions.]

Scoring

[Scoring for each design should be recorded using the participant's score sheet. The following information should be entered for each trial/design for each participant:

- Number of seconds required to complete the design
- Accuracy of the construction (Y or N).

Scores are calculated as follows:

- Participants receive 2 points each for completing Designs 1 through 4 on the first trial (8 points possible).

- Participants receive 1 point each for completing Designs 1 through 4 on the second trial.
- If a participant completes Designs 3 and 4 on the first trial, 4 points are automatically awarded for Designs 1 and 2. (Note: Designs 1 and 2 are only performed if the participant fails to complete design 3 or 4.)
- Participants receive 4 points each for completing Designs 5 through 9. The participant can earn from 1 to 3 bonus points for faster performance. The table below shows the total points awarded based on completion time. (Note: Designs 10-13 are not relevant for this experiment. In this study, participants will only be required to perform block Designs 5 through 9 in addition to diagnostic Design 3 and 4.)

			Points awarded for completion times			
#	Blocks	Time limit	4	5	6	7
5	4	60	21-60	16-20	11-20	1-10
6	4	60	21-60	16-20	11-20	1-10
7	4	60	21-60	16-20	11-20	1-10
8	4	60	21-60	16-20	11-20	1-10
9	4	60	21-60	16-20	11-20	1-10
10	9	120	66-120	46-65	31-45	1-30
11	9	120	76-120	56-75	41-55	1-40
12	9	120	76-120	56-75	41-55	1-40
13	9	120	76-120	56-75	41-55	1-40

The participant must not be allowed to see the score sheet during testing.]

VR Materials

[For the **VR task**, the VR BD workstation is used instead of the stimulus booklet. The materials necessary for this testing include:

1. A stopwatch
2. Multiple score sheets. Each score sheet should be identified as PXX_TY_ZZ, where XX is the participant number, Y is the session number, ZZ is the trial number.
3. NASA-Task Load Index (TLX) demand component ranking and rating sheets. Each sheet should be labeled in the same manner as the BD task score sheets.

The following system settings should also be verified for the testing:

1. Monitors set to 800/600 resolution.
2. Monitor refresh rate set at 120Hz.
3. A “\Results” directory should exist on the hard drive of the computer workstation.]

VR Procedure

[The participant will complete one BD test (i.e., one trial). An initial training trial should be provided. The test should follow the standard WASI sequence.
Similar to native BD procedure, participants should only be required to reproduce block designs 4 through 9 in addition to the diagnostic Designs 3.]

Script for Native Task (Native WASI rules)

<table>
<tr><td>Introductio
n</td><td>[Help the participant to adjust the chair height and backrest angle to sit comfortably. Mark the chair position with erasable marker pen and make sure the chair position is the same throughout the test. Apply the goniometer to make sure the participant achieve the following requirements:
Both elbows are rested on the test table while the upper arms are vertical with the table surface.
Shanks of the participant are vertical with the floor (use foot rest if necessary).
The head of the participant has an angle of 70 degrees with horizontal line when looking at the monitor.
Ask the participant to keep the trunk still while performing the task.]

Your task will be to reproduce a set of block designs using a set of 4 blocks. The blocks need to be placed according to the colored grid squares in the stimulus booklet.

[Confirm the overhead camera is vertically above and fixed on the entire work surface (i.e., the entire embedded monitor is visible). The video recording will be used for identifying the fine tuning phase of the assembly task and marking the EMG file for analysis purposes.

Start a new EMG file for each new design.]</td></tr>
<tr><td>Design 3</td><td>[Design 3 is for the first trial only and will not be included in subsequent trials or sessions.

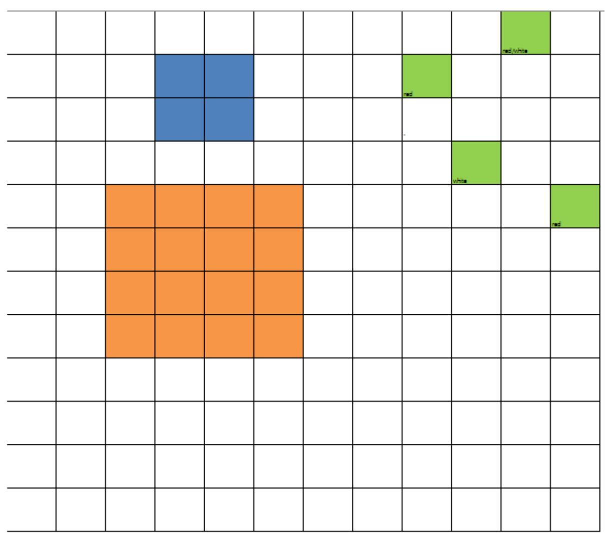
Place 4 blocks in the green squares on the monitor. Only one combination red and white side should be face up, the other three should be two red sides and one white side.]

Here you see the native BD task layout. There are four construction blocks positioned at random positions in the far right corner of the workspace. You can also see the design construction space towards the far left corner of the space. The orange square is the location in the workspace over which all block rotations must occur. I will now put the construction blocks together to make a design. Please watch me.</td></tr>
</table>

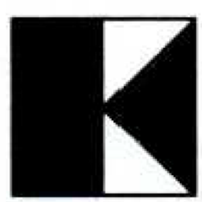
[Construct the design pictured below.]

You will move each block from the starting point to the orange square. Do not rotate the block until the block is over the orange square. After the block is directly over the orange square, rotate it to the proper orientation. After rotating the block, place it in the blue square to construct the design. For every block, even if rotation is not needed, you must move it into the orange square before putting it into the block square. While performing this task, you can drop the block on table whenever needed; however, to adjust the block position, you must pick it up instead of pushing it.

Do you have any questions on this procedure?



[Do not explain the design itself. Leave the model intact 7-inches from the edge of the table in front of the participant.]



[Repeat instructions to participant.] Once again, *you should only rotate the block when it's over the orange square. I picked-up the blocks and moved them to the orange grid before rotation. Once over the orange grid, I oriented a block to a particular side. Next, I moved the block to the blue target grid and placed the block in order to create the correct design. You*

<table>
<tr><td></td><td> should follow these same steps in reproducing the designs.
Once in the blue grid you cannot rotate the block. However, you can move the block back to the orange grid to re-orient the block.
[Move the assembled model you just made of off the surface of the monitor, but in view of the participant. The participant will need to be able to see it. Place the remaining four (4) blocks back in the green starting positions. Only one cross hatched side should be face up, the other three blocks should be two reds and one white.]
[Make sure the participant's sitting position is in accordance with prior requirements with a goniometer.]
Now, I ask that you create a block pattern just like this (point to the model). Try to work as fast as possible. Remember to rotate the blocks only over the orange square. Also, while completing the task, please keep your sitting position still.
[Start recording the EMG data, then start the timer. Watch the monitor for the overhead camera. Mark the EMG file every time a block enters and exits the orange square by pressing "Esc" button on the keyboard.
Stop the EMG file when the participant completes the design.
Provide another two trials of Design 3 for practice. Ask the participant to get familiar with the task requirements and sitting position while performing the task.] </td></tr>
<tr><td>NASA-
TLX</td><td> [Present the participant with the NASA-TLX demand component rank form.]
I now ask that you complete a pairwise ranking of task workload dimensions based on your training experience. On this sheet, you see pairs of workload dimensions listed on each line. Given your experience with completing the native BD task, please circle the demand component on each line of the sheet that you feel was more important to task performance. Do you have any questions on the use of this sheet? </td></tr>
<tr><td>Design
4-9

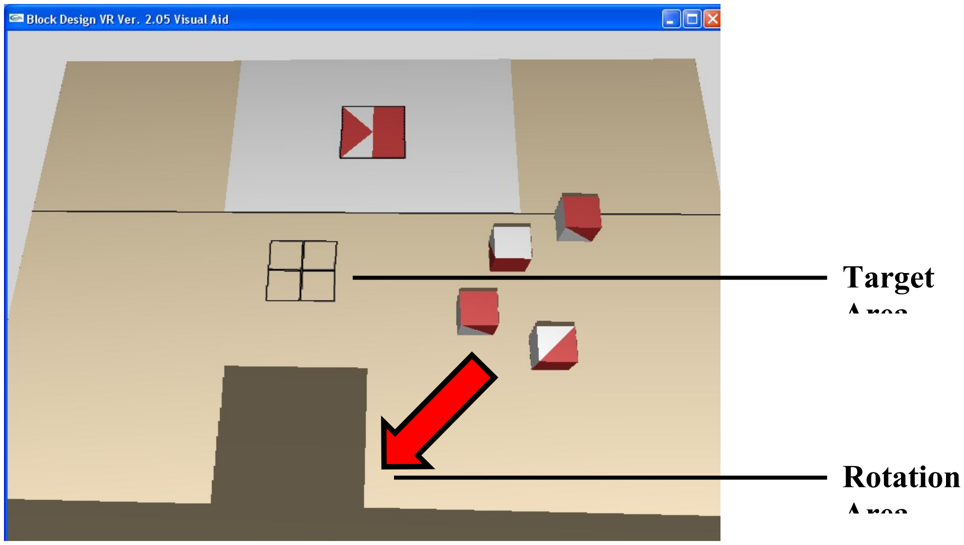
(60 sec)</td><td> [Use the test booklet to present the designs.
Place the four (4) blocks back in the green starting positions. Only one cross hatched side should be face up. The other three blocks should be two reds </td></tr>
</table>

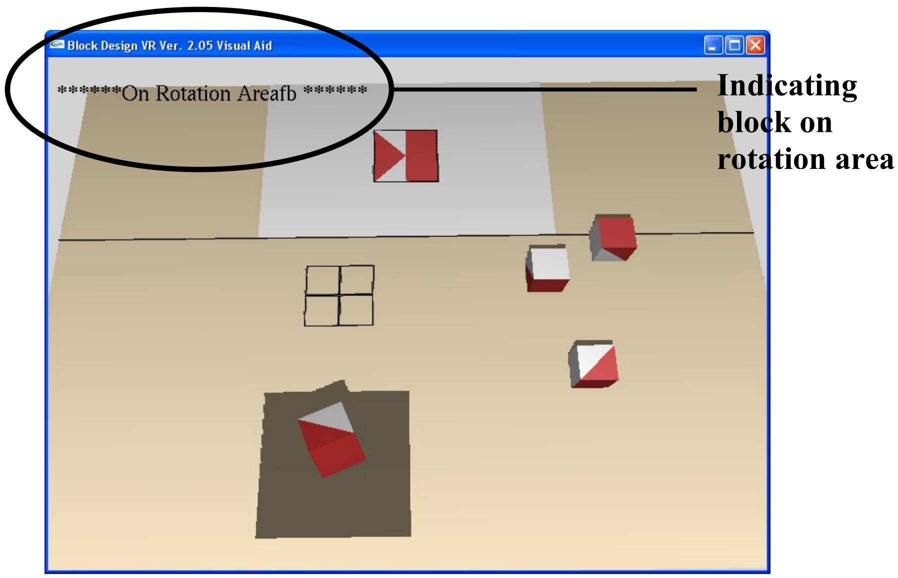
	and one white. Turn to the next page of the test booklet to expose the next design.] [Before starting every trial, use the goniometer to make sure the sitting position meets requirements.] *Now, please create a design pattern just like this one* (point to the pattern). *Try to work as fast as possible. Remember to only rotate the blocks over the orange square and every block has to be moved through the orange square even if it does not need rotation. To adjust the position of any block on table, you have to pick it up and drop down again. Do not push any blocks to change their positions. Also, please keep your sitting position still while completing the task.* *Tell me when you are finished. Go ahead and start.* **[Start recording the EMG data, then start the timer**. Watch the monitor for the overhead camera. Mark the EMG file every time a block **enters** or **exits** the orange square by pressing "Esc" button on the keyboard. Stop the EMG file recording when the participant completes the design. Save the file Repeat Design 4-9.]

Script for VR Task

Motor skill training	*I will now show you a demonstration for the VR version of the BD test.* *In this test, you will perform a block design task using the block-shaped control device as in previous training session.* [Ask the participant to put on the 3D goggles. Lift the block control before starting the trial.] *The function of the block control will be the same as in dice training tasks when manipulating the virtual blocks. During this training, your task is to construct designs using a set of 4 virtual blocks. As you can see, these blocks* [point to the blocks on the screen] *are identical. Some sides are whole white, some are whole red, and some are half and half. Please move the given blocks into the grids* [point to the grids on the virtual desk as desired destination] *to make the combined top surface look exactly like the given stimuli* [point to the stimuli design on the screen].

	Do you have any questions about the instructions? You will be given a chance to try the task before we begin recording your performance.
Device Training Design 3 (No time limits)	[Ask the participant to sit on the test chair with arm rest and face the test table with embedded monitor. Help them adjust the chair to achieve the following requirements: • Both elbows are rested on the test table while the upper arms are vertical with the table surface. • Shanks of the participant are vertical with the floor (use foot rest if necessary). Make sure the arm rest on right side is at the same height with the test table. Move the chair in front of the table with VR monitor. Mark the chair position with erasable marker pen and make sure the chair position is the same throughout the test.] [Start the block design training software on the VR workstation. Type in the participant number and trial number. Choose **Design 3.** Set the view angle as 0.35 radian (i.e., 20 degrees).] *Now, as you saw in the video, I am going to ask you to create the same design patterns using the block-shaped control.* [Point to the block control]. *You can grab the block by clicking the button.* [Show the position of button.] *While performing the task, you are not required to keep the same posture in holding the block control. It is OK to rotate the block control when necessary.* *This first design will be used to help acquaint you with the VR BD task. Make sure to pick up the block and move it to the large shadowed square. Once positioned in the large square, you can rotate the block in the correct manner. Next, you can move the block to the small target grid and position it in order to create the correct design. Once positioned in the target grid, you cannot re-orient a block. However, you can move the block back to the large grid for rotation. While performing the task, please keep your trunk still.*

	 *Do you have any questions? Are you ready to begin?* [Provide another two trials of Design 3 for practice. Ask the participant to get familiar with the task requirements and sitting position while performing the task.]
NASA-TLX	[Present the participant with the NASA-TLX demand component rank form.] *I now ask that you complete a pairwise ranking of task workload dimensions based on your training experience. On this sheet, you see pairs of workload dimensions listed on each line. Given your experience with the use of the haptic device in completing the VR BD task, please circle the demand component on each line of the sheet that you feel was more important to task performance. Do you have any questions on the use of this sheet?*
Design 4-9 (3 min per trial)	*Thank you for completing the workload demand ranking for the task. Now, we will go forward with the formal test session.* *We will record your task completion time from this point forward. A timer will start automatically when you see the scrambled blocks on the screen. I ask that you tell me when you finish a design pattern. You have at most 3 minutes to complete each design.* *Try to work as fast as possible. Remember to only rotate the blocks over the large square. While performing the task, please keep your trunk still.* [Once the participant completes a design pattern, enter the number of the next

	Design (4-9). Lift the block-control above the table surface. Give the block-control to the participant.] *Are you Ready? Remember to tell me as soon as you think you have completed reproducing a design pattern.* **[Start recording the EMG data and then start the trial by clicking 'OK' on the display screen**. Watch the upper left corner of the VR monitor for the message, "On the rotation area." Mark the EMG file every time the message appears **and** when it disappears by pressing "Esc" button on the keyboard. Stop the EMG file recording when the participant completes the design. Save the file.]  *You have now completed the current block design. Please relax for a moment while I reset the system and present the next design pattern.* [Repeat the above procedure for tests 4-9.]

Departure
[The following should be read at the end of the final session.]

NASA TLX	[Present the participant with the NASA-TLX demand component rating form.] *Please fill-out this workload rating form in order to assess the difficulty of the task. We ask that you make a rating for each workload dimension by writing an "x" on each rating scale between "low" and "high". Do you have any*

	questions on the use of this form?
EMG Electrode Removal	*Thank you for completing the workload demand rating form.* [Turn all EMG equipment off (Biopac amplifiers and system).] *Now the experiment session is complete. I will begin removing all electrodes and tape. Slight redness or irritation of your skin may occur after removal of the electrodes. This reaction is normal and the symptoms will subside in a few minutes.* • [Begin CAREFULLY removing electrodes off a participant's forearm according to the following steps: Remove the adhesive tape and disconnect the leads from the electrodes. • Disconnect the other end of the lead from the socket in EMG100c module. • Remove and discard the electrodes from the participant's skin. • Clean any residual skin prep gel from the participant's skin by gently wiping with rubbing alcohol and a cotton ball. Do not let a participant remove the electrodes. Do not pull wires to remove the electrodes.]
Departure and thanks	[Calculate the participant payment based on the time spent in the lab.] *Please fill-in this payment form. The compensation for you today is $xx.* [Let the participant complete the payment form.] *The data collected today will be used to determine whether the design of the VR simulation, integrating the haptic control interface, is suitable for motor rehabilitation and skill training.* *If you are interested in future information about this study or have any questions, please contact Dr. David Kaber. His contact information is listed in the consent form that you signed.* [Give the participant a copy of the consent form and direct them to see Hakan Sungur for payment.]

	Here is a copy of the informed consent form that you signed at the beginning of the experiment. Please take your payment form to Mr. Hakan Sungur in Room 423 in order to receive your payment. *Thank you for participating in this study.*