COGNITIVE ABILITY AND COMPUTER SIMULATION

By

Charles Colby Buford

Brian J. O'Leary
Department Head - Psychology
(Chair)

Christopher J. L. Cunningham
UC Foundation Associate Professor of Psychology
(Committee Member)

Michael D. Biderman
Salem Carpet Professor of Psychology
(Committee Member)

Joe Dumas
UC Foundation Professor of Computer Science
(Committee Member)

COGNITIVE ABILITY AND COMPUTER SIMULATION

By

Charles Colby Buford

A Thesis Submitted to the Faculty of the University of
Tennessee at Chattanooga in Partial Fulfillment of the
Requirements of the Degree of Master's of
Science: Psychology

The University of Tennessee at Chattanooga
Chattanooga, Tennessee

May 2014

ABSTRACT

General Mental Ability (GMA) is the single strongest predictor of future work performance currently available. Tests of GMA used for selection purposes carry potential for adverse impact if used as a primary tool for employee selection; however, a factor of GMA known as Fluid Intelligence (Gf) should represent a more equitable assessment construct. Computer simulations appear to offer some beneficial attributes for selection purposes, but potential negative effects of GMA assessment may carry over to the new medium. In this study, I successfully used a computer simulated game to approximate two tests of Gf with a study of both pilot and in-person samples. Results indicated that a person's prior experience and skill with computer simulated games does not influence predictions of Gf using a simulated game.

DEDICATION

This work is dedicated to all of my family, whom have provided me such strong support for me to build a life upon, this one's for you. I love you all.

ACKNOWLEDGEMENTS

This work would not have been possible if not for the researchers who have expended a tremendous amount of effort and time over the years to teach me the principles and methods of ethical research. To Dr.'s Brian O'Leary, Eric Seemann, Christopher Cunningham, Jodi Price, Mike Biderman, and Aurora Torres; your time with me has made a difference. To my friends who have enriched my life so much, I hope that I've been able to enrich yours half as well.

A project like this takes a great deal of work from many sources and attempts to combine them to bring out the best of our discipline's knowledge. The research programs offered by Wonderlic, Pearson, and Western Psychological Services have gone a long way towards making this study financially possible. UTC's Provost Research Award provided the funding used to purchase the tools and equipment needed for this project. Without the research discounts from these publishers and funding through UTC, this research could not have taken place. Thank you for your support!

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

Gc, Crystallized Intelligence

Gf, Fluid Intelligence

GMA, General Mental Ability

g, General Intelligence

RPM, Raven's Progressive Matrices

SD, Standard Deviation

SE, Standard Error

M, Mean

LIST OF SYMBOLS

r, Pearson product-moment correlation coefficient, a measure of the strength and direction of the linear relationship between two variables.

t, T-test between two independent samples, a measure of the difference between two mean scores from different participant samples.

p, Probability value, a probability of obtaining a test statistic value of at least the same value if computed using another sample; a significant value indicates that a measured value was unlikely to occur again by chance alone.

b, Unstandardized Beta weight, within a linear regression, represents the change in the dependent variable that is associated with a change in one standard deviation of the independent variable when holding other entered predictors constant.

β, Standardized Beta weight, represents the same concepts as an Unstandardized Beta weight, but each predictor is transformed into a standardized value allowing comparison between different or unfamiliar scales.

R^2, R-Squared, the percentage of a dependent variable predicted by all predictors entered into a regression model.

ΔR^2, Delta R-Squared, the change in R^2 caused by the addition or removal of a regression model's predictors.

CHAPTER I

INTRODUCTION AND LITERATURE REVIEW

Overview

There are many selection methods intended to identify high performing employees. Research suggests that selection procedures based on General Mental Ability (GMA) are the single best predictor of later job performance (Schmidt & Hunter, 1998). Selection measures based on GMA are useful for a variety of job types, but bring with them practical and ethical problems, including the potential for creating adverse impact for women and minorities in selection and other employment-related decisions. Several attempts to resolve these problems have failed (Brody, 2007; J. E. Hunter, 1986; Naglieri & Jensen, 1987). Newer methods, including computer simulated assessment, may offer new opportunities to maintain strong performance prediction while overcoming adverse impact. Measures that utilize non-verbal (Fluid) intelligence over GMA or verbal (Crystallized) intelligence should be able to overcome adverse impact effects by drawing on natural abilities rather than cultural or educational knowledge. In the current study, a computer simulation was modified from an existing commercial computer game to function as a measure of non-verbal (Fluid) intelligence.

Cognitive Ability

Research on Macon be traced to the work of Cattell, Horn, and Carroll, which has been synthesized into the Cattell-Horn-Carroll (CHC) theory of intelligence (McGrew, 2009). CHC theory describes three strata of cognitive ability consisting of: 1) a general intelligence factor (*g*) at the highest level, 2) several broad abilities, and 3) many specific and narrow abilities (Carroll, 1997; McGrew, 2009). Specific and narrow abilities are used to measure broad abilities and form the basis for intelligence test scale creation (Carroll, 1997). The general intelligence factor *g* represents the degree to which a person can process as well as learn information, make decisions, and reason their actions, and these capacities are known to improve a person's ability to perform any task when other factors are held constant (Horn, 1968; Horn &Cattell, 1966; Schmidt & Hunter, 1998). Most used of the broad, second-level abilities are crystallized intelligence (Gc) and fluid intelligence (Gf) (Johnson & Bouchardjr, 2005). Gc is the consolidated knowledge a person accumulates through education, culture, and instruction and its development is influenced by Gf (Horn & Cattell, 1966; Johnson & Bouchardjr, 2005; McGrew, 2009). Gf is the ability to perceive relationships, reason, and consider abstractions; it is not influenced by prior learning or acculturation, but is determined by incidental or uninstructed learning (Horn & Cattell, 1966; Horn, 1968). Gf is thought to be fairer to those of lower socioeconomic status or from different cultures due to its reliance on abilities that are not influenced by social status (e.g., education, socialization). Working memory is another second level ability that closely resembles Gf; the combination of these two factors can form an approximation of the ability to maintain attention and response patterns in the face of distractions (Engle, Tuholski, Laughlin, & Conway, 1999).

Selection of Employees

GMA is directly linked to job performance across a wide variety of job types, job complexities, and cultural differences (Hunter, 1968; Hunter & Schmidt, 1996; Schmidt, 2002; Schmidt & Hunter, 1998), ranging from telemarketers to police officers, and airplane or fighter pilots (Carretta, 2011; Gordon & Leighty, 1988; Hakstian, Scratchley, MacLeod, Tweed, & Siddarth, 1997; Pynes & Bernardin, 1989). GMA also predicts interview performance (r = .40), providing a measure of both future performance and the likelihood of being hired under a typical employment battery (S. T. Hunter, Cushenbery, & Friedrich, 2012; Roth & Huffcutt, 2013).

Gf is the component of GMA that is associated with the ability to think divergently, creatively, and to multitask effectively (Batey, Furnham, & Safiullina, 2010; S. T. Hunter et al., 2012; Konig, Buhner, & Murling, 2005; Roper & Juneja, 2008). Research indicates that the information-processing and decision making demands of a job, load heavily onto Gf abilities and serve as a meaningful moderator for the GMA and job performance relationship (Gutenberg, Arvey, Osburn, & Jeanneret, 1983).

Research consistently finds that possessing a higher GMA increases a person's ability to solve problems that are both highly specific and general in nature (Beier & Oswald, 2012; Osato & Sherry, 1993; Ree & Earles, 1992). Tests of GMA have also been found to be equivalent when taken on a computer or with more traditional paper-based methods (Mead & Drasgow, 1993). However, GMA tests do not display the same predictive ability when a person is retested using the same test, instead measuring primarily memory, which may cause issues with applicants applying to multiple job positions using the same off-the-shelf test (Lievens, Reeve, & Heggestad, 2007).

Subgroup Differences

Despite the strong positive correlation between GMA and job performance, GMA test scores across a wide variety of studies have consistently demonstrated differences by ethnicity and sex(Brody, 2007; Gottfredson, 2000). The largest and most consistently identified gap is the one standard deviation difference between Caucasians and African Americans, with African Americans scoring lower than Caucasians (Gottfredson, 2000; Roth, Bevier, Bobko, Switzer, & Tyler, 2001). While many efforts have been made to close this gap, the difference is stubborn and some researchers have seemingly resigned themselves to accepting differences in scores as representative of a true difference of ability for African Americans (Gottfredson, 2000; Rushton & Jensen, 2005; Schmidt, 2002).

Addressing Score Differences

Attempts to reduce the adverse impact of GMA measures by modifying test content have not improved the efficiency or effectiveness of the tests (Duckworth, Quinn, Lynam, Loeber, & Stouthamer-Loeber, 2011; Ployhart & Ehrhart, 2002). Focusing solely on specific components of GMA in traditional measures, such as Gf or Gc, rather than *g*, does not result in a reduction in adverse impact without a corresponding or greater reduction in predictive ability (Waters, 2007). It seems that, when changes in test content are successful in reducing adverse impact for one subgroup (e.g., ethnicity), another subgroup (e.g. sex) is negatively affected (Sager, Peterson, Oppler, Rosse, & Walker, 1997). Measuring job-specific cognitive abilities by creating tailored assessments is also problematic as these measures often carryover the adverse impact effects of GMA (Naglieri & Jensen, 1987).Further, research indicates that non-cognitive contextual factors, such as using traditional paper test administration, affects GMA test scores by increasing anxiety, reducing motivation, and inducing stereotype threat in African Americans (McKay & Doverspike, 2001).

The difficulties of using traditional measures of GMA and its specific predictors have led many researchers to attempt the use of alternative predictors paired with tests of GMA. Schmidt and Hunter (1998) suggested combining other assessments, such as integrity tests and structured interviews, with GMA to maximize predictive ability. Unfortunately, their top recommendations are less than ideal, as integrity tests are only minimally predictive of future performance, and structured interviews share many of GMA's predictive and adverse properties (S. T. Hunter et al., 2012; Roth & Huffcutt, 2013; Van Iddekinge, Roth, Raymark, & Odle-Dusseau, 2012).

Computer Simulations as Assessments

Computer simulations are a technology that have only relatively recently been used as training and evaluation tools for both military and civilian needs. Learners using these tools are given opportunities to practice their skills during training, helping them to solidify learned behaviors before applying them to a job (Jong, 1991; Ortiz, 1994). Simulations used for training have demonstrated equal effectiveness when designed as either strict simulation of real environments or as a simulated game (Jentsch & Bowers, 1998). Research has found that simulations should be designed to encourage participants to view errors as opportunities to learn, as interactive activities, such as simulations, can lead to frustration (Bell & Kozlowski, 2008; Hughes et al., 2013).

Using computer simulations to assess intelligence is a relatively untested area. However, inductive reasoning (a component of Gf), as measured by the Berlin Model of Intelligence Structure, has been successfully measured with a simulation, with both assessments demonstrating a relationship strength that is similar to those found between two traditional tests of intelligence ($r = .75$) (Kroner, Plass, & Leutner, 2005). Hughes et al. (2013) noted that these relationships require controlling for a person's prior experience and exposure to similar technologies, other simulations, and games to accurately assess these relationships. However, Mennecke, Hassall, and Triplett (2008) reported that performance within a simulated environment requires a period of acclimation as respondents learn how to interact with a simulation.

Simulations and games are perceived by industry as legitimate alternatives to traditional assessment methods by applicants, investors, and supervisors (Fetzer & Tuzinski, 2013). Some criterion studies have shown correlations between job performance metrics and simulated assessments above.40 for customer service, sales, teller operator, and collection agents (Boyce, Corbet, & Adler, 2013). Candidates reported that simulations provided a fair assessment offering sufficient opportunity to perform (Boyce et al., 2013). The relatively high fidelity of simulations with the actual work content and environment may also provide a realistic job preview, and are viewed favorably by internal stakeholders (Boyce et al., 2013).

Despite a recent surge in applied simulation use, many of these simulations remain proprietary, preventing others from using them to perform additional, independent research. This is unfortunate, as simulations may offer a potential measure of nonverbal intelligence because simulated game performance is primarily linked to the perceptual and cognitive abilities in the Gf domain (Richardson, Powers, & Bousquet, 2011). Existing research suggests that performance on simulations can be used to approximate a work sample and predict some degree of future job performance (Motowidlo, Dunnette, & Carter, 1990). Job candidates also typically perceive multimedia simulations as more face valid, job related, and engaging than traditional selection methods (Chan & Schmitt, 2004; Mayer, 2004; Motowidlo et al., 1990; Richman-Hirsch, 2000). The ability of simulations to approximate measures of job performance raises the question as to how simulations relate to GMA, another strong predictor of work performance (Schmidt & Hunter, 1998).

The Present Study

The use of simulations as measures of GMA remains relatively unexplored. To begin to address this gap in the literature, I conducted a validation study using a simulation as a measure of fluid intelligence (Gf). I modified an existing computer game, Portal 2, to function as a structured assessment. I compared performance on the modified simulation with existing measures of GMA and its Gf and Gc components, using the Wonderlic, Raven's Progressive Matrices (RPM) and the Shipley-2 intelligence tests.

Existing Measure of GMA

I chose the tests used in this study for their unique contributions toward understanding the nature of simulation performance for intelligence assessment. The Shipley-2 stood out among these measures by providing an assessment of both Gf and Gc abilities within the same published test. Its Block Patterns form appeared to supply an assessment of Deductive reasoning by having respondents observe a presented shape and to determine what shapes would cause a response picture to match the question by applying a similarity rule. The RPM provided a closely related assessment of Inductive reasoning by asking participants to perceive a pattern within a response question and respond with the missing value. The Shipley-2's Vocabulary form served as a measure of lexical knowledge, one of the core factors of Gc, by examining the extent of participant knowledge of word meanings and concepts (Schneider & McGrew, 2012). The Wonderlic served as a general intelligence measure by including multiple different abilities for a general assessment of *g* across many domains of ability. Through the application of these tests, I hoped to detect both Inductive and Deductive reasoning with two separate measures, compared to another specific ability within the Gc factor, as well as a general assessment of ability.

Portal 2

Portal 2 is a computer game first published in 2011, in which players control a character using a first person viewpoint to solve a series of room-based puzzles. These puzzles are environmental in nature and typically focus on moving the character from one location to another by activating devices. Correctly completing a game level requires a player to observe all of the room's characteristics, such as puzzle elements, the room's shape, and how devices are activated, reason a potential solution to the room, and successfully execute the solution. This process appears to be closely aligned with Induction and Deductive

Reasoning which are two specific and narrow abilities included in the CHC model's Gf factor (Carroll, 1997; McGrew, 2009). Induction has been described as the ability to observe and discover patterns or rules(Schneider & McGrew, 2012). Deductive Reasoning represents the capacity to logically reason using known information or rules (Schneider & McGrew, 2012). These narrow abilities have previously been demonstrated to best reflect Gf and I believed that the process of playing the Portal 2 game would represent an expression of these abilities (McGrew, 2009; Schneider & McGrew, 2012).

The original version of Portal 2 includes several gameplay devices that may cause player death or which share some degree of redundancy to other puzzle elements. Player death mechanics were removed to prevent any frustration or anxiety that could potentially be caused by an instant failure and repeat of a level, while redundant devices were removed for time considerations. A total of five puzzle elements were retained in this modified version to maximize differences between puzzle elements; a full description of these devices is included in the Materials section below.

Portal 2 uses a built-in console command that automatically records player performance for each completed level. Performance is recorded after completing a level and includes the elapsed time to complete each level, the number of portals (i.e., doorways) created for each level, the number of steps taken by the character within each level, the number of levels completed, and the number of deaths within each level. The player death metric was removed from the modified version and is not included in later sections.

I believed that each of the three performance metrics would represent a particular aspect of performance within the game and would each contribute toward the Gf construct. The elapsed time per level indicated how quickly a person completed a puzzle, measuring an overall score of a participant's inductive and deductive reasoning attempts for that level. The number of portals created per level represented the efficiency of a participant's inductive and deductive reasoning more attributable to the characteristics of each level, as each portal represents a manipulation to the puzzle. The number of steps taken by the character per level provided a supplemental measure of performance indicating deliberation by a participant as they paced around each puzzle to rotate their character's viewpoint. The total number of levels completed indicated how many successful solutions the participant could generate within the test period, measuring an overall score of reasoning similar to time performance, but for the entire play session.

Hypotheses

Tests utilizing the CHC model use a theoretical model that assesses specific and narrow abilities combined into factors (Gf, Gc) which are then combined into to an overall score (*g*). Completing each game level requires participants to express a degree of Induction and Deductive Reasoning; if the expression of these abilities is of a sufficient magnitude within the game's performance metrics, game performance should reflect Gf. Therefore my first hypothesis follows:

(H1) A participants' performance on the Portal 2 Gf simulation is positively correlated with Gf.

The predictive ability of prior experience on simulation performance has been consistently found in the simulation literature (Boyce et al., 2013; Hughes et al., 2013). Therefore I predicted my second hypothesis:

(H2) Previous game experience is positively related to performance within the Portal 2 Gf simulation.

Prior experience has also been shown to moderate relationships between simulation performance and other external measures or outcomes, in addition to its main effect on game performance, demonstrating a strong need to account for prior game experience (Boyce et al., 2013; Hughes et al., 2013).Therefore:

(H3) Game experience moderates the positive relationship between Portal 2 Gf simulation performance and Gf, such that greater experience strengthens the relationship.

CHAPTER II

METHOD

Participants

Two pools of participants were recruited: one using an online pilot and the other in-person. The data collection process is summarized by sample below.

Pilot Sample

An online pilot sample (*N* = 94) was collected through an internet community website (www.somethingawful.com), which was populated by many users familiar with and experienced using Portal 2 and similar games. Participants were included in a$100 random lottery drawing for completing the modified Portal 2 game and emailing completed performance data. Participants were later offered an additional $100 lottery drawing to return after several weeks to complete the Wonderlic online test if they had previously completed the simulated game. This returning sample was smaller than expected (*n* = 27) but is still included in results for comprehensiveness.

In-Person Sample

The in-person undergraduate sample (*n* = 73) was collected from a medium-sized, southeastern university. Participants were offered research extra credit for use in their undergraduate psychology courses as well as entry in a $100 random lottery drawing for completing the RPM, Shipley-2, and the Portal 2 Gf game.

Descriptive Profile

Summary descriptive statistics for both samples are presented in Appendix A, Table 1. Pilot participants (*n*= 94) included 85 (90.4%) reporting as Caucasian, with three (3.2%) Hispanic, two (2.1%)

Asian, two (2.1%) Inter-racial, and two (2.1%) other ethnicities. The sample was predominantly male (*n* = 87, 92.6%) creating a predominantly Caucasian male sample. Notably, this sample reported a higher mean age and a higher degree of gameplay experience than the in-person sample (Appendix B, Figure 6).

The in-person sample (*n* = 73) consisted of a total of 53 (72.6%) participants reported as Caucasian, with nine (12.3%) African American, five (6.8%) Hispanic, four (5.5%) Inter-Racial, and two (2.7%) Asian. The sample included 42 women (57.5%) that, while still predominantly Caucasian, provided a more balanced gender ratio than in the pilot sample.

Materials

Participants (both samples) completed an online demographic survey including first and last name, participant and computer terminal number (if in person), preferred method of contact, current age, sex, and ethnicity. Game experience questions determined how often a participant played electronic (video) games (ranging from once per day to not normally playing with six possible choices), as well as self-reported skill (Skilled, Average, Unskilled) at electronic games, electronic puzzle games, and experience with Portal or Portal 2 (Skilled, Average, Unskilled, Have not played Portal/Portal 2). This demographic survey is presented in Appendix C. A gameplay experience score was created by summing on a numeric scale of one to six (for play frequency), one to three (for play skill), and one to four (for Portal experience), to create an experience score that ranged from four (very low) to sixteen (very high).

Sixteen temporary Portal 2 game licenses were obtained by submitting an application for educational use using a website form (www.teachwithportals.com) offered by the publisher of the simulated game, Valve Corporation. I created game levels using both a basic editor within the game, *Portal 2 Puzzle Maker*, as well as an advanced external editing program, *Valve Hammer Editor*, each of which was provided free of charge with the Portal 2 license. I created a series of unique, custom game levels for the current study with the purpose of preventing any possibility of prior exposure to a similar, preexisting game level. I thus ensured that, while participants varied as a function of their prior exposure, all game levels were new experiences for all participants. Twenty-six levels were created for the study, and were arranged into three series to acclimate and assess participants with the simulation.

Consistent with recommendations by Mennecke et al. (2008), I developed seven initial training levels to teach the five puzzle elements and the Portal 2 game, with audio narration, first by introducing these game elements alone within each level, and then combining them with other previously taught devices to gradually introduce more complex gameplay puzzle elements and their combinations with other puzzle elements. Training levels and puzzles were constructed to be linear to limit the potential for a player to attempt incorrect solutions while learning the game and to focus narration on the participant's relative position in the game level. The first game level focused on orienting the player to the control scheme of the game and had participants walk through a short maze and place a cube onto a button power the exit door. The second and fourth training levels focused on initially using the portal gun to create portals to move across distances and then later introduced them as a method of transferring inertia of objects that were falling (Appendix B, Figure 2). These portals allowed participants to place a pair of spatially-linked portals (or doorways) that could be used to cross distances, reorient game devices, and allow for unusual ways of movement. A third training level introduced the faith plates puzzle device, a device which acted as a catapult for the player and game objects. The fifth training level introduced lasers, which powered other game devices when they were redirected using portals or a special cube (Appendix B, Figure 3). A sixth training level introduced a blue goo that, when redirected from a dispenser, could visually coat surfaces in blue and transfer a reflexive property to them; they would then act similarly to a trampoline when touched (Appendix B, Figure 3). The final training level introduced the excursion funnel, which acted as a continuous pushing force in one direction that could move the player, blue goo, or cubes (Appendix B, Figure 4). A full list of these puzzle elements are located in Appendix A, Table 1.

Having introduced all relevant game concepts, four levels were created to evaluate the effectiveness of the training. These evaluation levels allowed players the freedom to attempt incorrect answers while continuing to maintain the level of puzzle complexity found within the training levels. The first evaluation level tested participants' ability to use momentum transferred with portals and redirecting a laser. The second evaluation level had participants use momentum transfer using portals and excursion funnels to cross distances. The third evaluation level required blue goo to be used to climb a room and to use portals to cross distances. The final training evaluation level had participants use blue goo and the excursion funnel individually to cross distances, move an object, and redirect a laser.

A final series of fifteen levels were created to evaluate participants by incrementally raising the number of required manipulations to each puzzle to correctly solve it. These levels used as many of the previously introduced puzzle elements as needed to incrementally increase the difficulty of each level from the last.

Audio narration recorded by a native English speaking female using a desktop microphone volunteer instructed and guided participants in gameplay concepts and through game levels. Audio was post-processed for clarity and volume using the Audacity 2.0.3 audio editing program. A computerized beep was added before and after narration audio clips to ensure that pauses in speech were not interpreted as breaks. Dialog was triggered by the player's position within the game, allowing planned instruction and guidance as well as allowing the player to continue moving within the game while narration played. Special indicated areas within each level (that featured narration) allowed participants to replay any audio narration they wished to hear again.

Procedure

The study was approved through the University of Tennessee at Chattanooga's Institutional Review Board. Pilot participants were required to have a personal copy of the Portal 2 game. They were given instructions and online technical support for extracting custom computer files to modify their personal copies of Portal 2 to run the modified simulation and enable performance recording procedures. Participants completed an online demographic survey, identical to both samples (Appendix C), and then played the modified Portal 2 game (Portal 2 Gf) for a maximum of 90 minutes before exiting the game and forwarding gameplay data through email. Several weeks after completion of this phase, pilot participants were asked to complete the Wonderlic online test.

In-person participants were read a verbal prompt detailing the study and its tasks, duration, and potential for nausea/motion sickness. Participants were asked to complete an informed consent document before being administered the Raven's Standard Progressive Matrices (RPM) test of Gf with a maximum given length of forty-five minutes. After a five minute break, participants completed the Shipley-2 Vocabulary form and Block Patterns form with ten minutes given to complete each form. They then received a fifteen minute break with refreshments after which they completed the online demographic survey.

Participants were then read a short prompt and shown a brief instructional video describing the nature, objective, and controls of the Portal 2 Gf game before playing for a period of 90 minutes. Participants wore headphones while playing to attend to game narration and sound effects that indicated player actions and puzzle elements. A projected image (Appendix B, Figure 3) remained in front of participants during this phase, reminding them of the controls of the game until the end of the 90 minute period. Participants were thanked and debriefed following this period.

Measures

Raven's Standard Progressive Matrices (RPM): The RPM is available in revised formats with three versions suitable for either children or impaired individuals, individuals of above-average intelligence, and individuals of normal intelligence (Raven, Raven, & Court, 2003). The tests are administered over a period of 45 minutes, and can be conducted in group settings for educational or organizational use to provide a measure of Gf (Raven et al., 2003). The test includes sixty questions presented in five groups of twelve items and focuses on nonverbal content, testing observation skills, learning ability, and problem solving (Raven et al., 2003). For each item, participants choose a fill-in-the-blank answer choice to a visual puzzle that contains a missing pattern or sequence. The first two groups of test questions present a total of six answer choices, and the remaining three groups present eight answer choices. Numerous studies have used the RPM and the test manual for the assessment lists it’s a split-half reliability of (.91) for young adults (Raven et al., 2003).

Wonderlic Personnel Quicktest: The Wonderlic has been well documented as a test of GMA in industrial and other workplace environments (Chan, 1997; Dodrill & Warner, 1988). I used the online version of the Wonderlic Personnel Test - Quicktest (or Pretest), which was electronically administered over a period of eight minutes. Response options for the Wonderlic Quicktest vary widely by question, with some items assessing word meanings (3-5 choices) and other items assessing numerical manipulation (5 options). Questions switch in focus throughout the test and are not separated into subtests.

Shipley Institute of Living Scale-2: The Shipley-2 provides an estimate of a person's Gf and Gc abilities, as well as an estimate of overall *g*, by combining these two scores. The test's Gf and Gc components distinguish it from other intelligence measures, but researchers have yet to use it extensively. The test consists of two subtests, Vocabulary for Gc, and a choice between either Abstraction or Block Patterns for Gf. The publisher describes the Block Patterns subtest as a comparatively purer measure of

Gf, and this form was used to approximate the Shipley-2's Gf score (Shipley, Gruber, Martin, & Klein, 2012). Reported split-half reliabilities for young adults range from (.85-.90) for the Vocabulary form and (.93-.94) for the Block Patterns form. Participants were asked to find the same meaning for a presented word on the Vocabulary form with a choice of circling one of four possible response words.

The Block Patterns form presents a smaller geometric pattern to the left of a larger matching pattern that is missing at least one area. The number of missing areas on the larger pattern increases as the test continues, with the first four questions missing only one area, the next five questions have two areas missing, and the final three questions have four missing areas. Response options are indicated by a letter next to each large pattern, while actual responses must be made directly onto the larger figure using multiple choice bubbles. Responding to this test is highly similar to the RPM; participants respond with a missing geometric shape to match the provided picture.

Portal 2Gf: Player performance was recorded through an automatic procedure which saved play data after each completed level. This procedure recorded the time taken to solve each game level in seconds (Time), the number of portal pairs created within each level (Portals), and the number of steps taken within each level (Steps).The number of levels completed was also recorded for each participant. Data for uncompleted levels was not recorded; thus all performance data related to a successfully completed level. Each of the game performance metrics for each level (Time, Portals, Steps) were converted into a standardized score through a z-score conversion based on the entire sample. These scores were each inverted to reflect that 1) less time taken, 2) fewer portals made, and 3) fewer steps taken, indicated better game performance. The number of levels completed was also converted into a standardized score (using the same method as the above game performance metrics) but was not inverted. The composite Portal 2 Gf score was calculated using the mean of the standardized number of levels completed, and the standardized and inverted results for Time, Portals, and Steps. As there were no previous studies using this design to suggest the appropriate weight for each component, each was weighted equally in calculating the composite score.

I combined the two participant samples for computing game performance and game experience scores, unless stated otherwise. As only in-person participants completed the RPM and Shipley-2, and pilot participants completed the Wonderlic, a comparison of intelligence tests across samples was not possible in this study.

CHAPTER III

RESULTS

Descriptive statistics for all samples are presented in Appendix A, Table 1 and correlations between all tests and performance measures for combined samples are presented in Appendix A, Table 3. Any significance levels greater than (p > .05) reported below are described as not significant (ns).

While participants in the online sample progressed through the game levels as anticipated, participants in the in-person sample were unable to make satisfactory progress within the simulation to assess them using only the initially planned set of 15 evaluation levels. As Appendix B, Figure 6 illustrates, over half of the in-person sample did not progress farther than the first training evaluation level during the 90 minute administration period. Despite this lack of progression, each of the training levels increased in complexity, so performance approximated what was anticipated in the original assessment method. As a result, all game levels were included in the calculation of game performance for both samples.

The reliability of the Portal 2 Gf game was assessed using a split-half reliability analysis by separately scoring even and odd game levels, creating a separate performance composite for each, and entering the results into a Spearman-Brown Prophecy formula. This procedure resulted in a split-half reliability of .92, suggesting strong internal reliability. These results are nearly identical for reported split-half reliabilities found for the test manuals of the RPM (.91), Shipley-2 Vocabulary (.85-.90), Shipley-2 Block Patterns (.93-.94), and the Wonderlic Personnel Quicktest (.85-.91).

The two samples differed significantly on gameplay experience scores (t = 15.32, p < .001; Pilot: M = 13.70, SD = 2.36; In-Person: M = 7.47, SD = 2.79) as well as sex composition (Pilot: 92.6% Male, In-person: 57.5% Female). As there was little ethnic diversity in the sample, I restricted my examination of participant differences to sex within the in-person sample and game experience across both samples. The in-person sample had a significant difference for gameplay experience by sex (t = 7.73, p < .001), with men reporting higher experience (M = 9.65, SD = 2.37) than women (M = 5.86, SD = 1.82).

Hypothesis one stated that test scores from the RPM would be positively correlated with game performance scores. As expected, game composite performance did correlate significantly with scores on the RPM in the combined sample (r = .44, p < .001). For the in-person sample, this relationship was found to be comparable to the pilot sample for men (r = .49, p <.01), and slightly weaker for women (r = .39, p < .01) (Appendix B, Figure 7). These results support hypothesis one.

Hypothesis two stated that previous game experience would be positively related to game performance. Results showed that, in the combined sample, game experience correlated highly with the Portal 2Gf performance composite (r = .79, p < .001), thus supporting Hypothesis two. However, further analysis of game performance for the in-person sample by sex revealed that, while experience correlated with Portal 2 Gf performance for men(r = .50, p < .01), it did not for women (r = .16, ns).

Hypothesis three proposed that game experience moderates the positive relationship between Portal 2 Gf game performance and Gf, with greater experience strengthening the relationship. Multiple linear regressions were conducted to examine the influence of adding game experience as a moderator of the relationship between Portal 2 Gf composite score and tests of Gf (RPM, Shipley-2 Block Patterns). Results (Appendix A, Table 4) showed no significant change in the relationship between Portal 2 Gf scores and either the RPM or Shipley-2 Block Patterns Gf tests. Therefore, hypothesis three was not supported.

CHAPTER IV

Discussion, Limitations, and Conclusion

Discussion

In this study, I developed a computerized assessment based upon an existing commercial computer simulated game, Portal 2, and examined how performance on modified game, Portal 2 Gf, related to established tests of Gf, Gc, and *g*. Results indicated a significant positive relationship between Portal 2 Gf performance and tests of Gf, the RPM and Shipley-2's Block Patterns Form, providing support for my first hypothesis. This finding provides support for the use of computer simulated games to assess nonverbal cognitive abilities. I did not find a significant relationship for game performance with *g* as measured by the Wonderlic Personnel test, and a weaker relationship for Gc as measured by the Shipley-2 Vocabulary test than for tests of Gf (Appendix A, Table 3). These results suggest that game performance was primarily associated with nonverbal cognitive abilities rather than abilities pertaining to verbal or general mental abilities. These relationships indicate that performance on Portal 2 Gf is capable of both convergent and discriminant validity as a measure of nonverbal ability. As would be expected, performance on Portal 2 Gf had a significant relationship with prior exposure to and experience with simulated games (H2).

Gender influenced these relationships, with men reporting a higher degree of game experience (M = 9.65, SD = 2.37) than women (M = 5.86, SD = 1.82). Regrettably, sex differences across samples could not be examined due to the small number of females in the pilot (Pilot: 92.6% Men, In-person: 57.5% Women). Relationships between existing measures of Gf and Portal 2 Gf performance were stronger for men, but also exhibited weaker relationships for women (Appendix A, Table 5). This trend may provide partial evidence for the negative effects of specific measures of cognitive ability on subgroup scores (Waters, 2007). Unfortunately, these differences could not be examined further for ethnicity due to relatively homogeneous samples.

The most surprising, but possibly most important result, is the lack of support for any moderating or mediating effect of game experience on the relationship between Portal 2 Gf performance and intelligence test scores. This indicates that a person's prior experience and exposure to computer simulated games does not influence the measurement of Gf using Portal 2 Gf. This is not to say that prior experience does not affect game performance, as the strong correlation ($r = .79$, $p < .001$) between game performance and experience clearly indicates, but that, when using Portal 2 Gf to measure nonverbal intelligence, the effects of prior experience are not significant. This effect may explain the seeming incongruence between this study's finding and other published works involving simulations and prior experience that did not focus on assessing nonverbal ability (Boyce et al., 2013; Hughes et al., 2013).

Exploratory Analysis

I conducted additional analyses to further investigate the relationships seen in the study. As the strongest relationship between game experience and game performance metrics was for the number of levels completed ($r = .87$, $p < .001$), all four game experience questionnaire items were regressed simultaneously onto the number of levels completed. This was a highly predictive model ($r = .91$, $R^2 = .82$, $p < .001$), with only the frequency of play (ß = .33, $p < .001$) and self-rated skill at Portal/Portal 2 (ß = .70, $p < .001$) questions functioning as independent predictors when controlling for each game experience question. These items were weighted and averaged to create a refined measure of game experience. However, this refined measure did not alter previous significance values or conclusions regarding a lack of moderation between game performance and intelligence measures.

One of the core abilities of more intelligent people is the ability to learn (Carroll, 1997). Each level of the Portal 2 simulation was designed to be more difficult than the previous level and, as the game required participants to continually interact with puzzles and game elements, this may have been encouraged a continuous learning process that approximated intelligence in its fluid form. Despite promising initial evidence and some extant support for these conclusions, more confirmatory evidence is required to support this possibility (Mennecke et al., 2008; Richardson et al., 2011; Richman-Hirsch, 2000).

Limitations

There are several limitations to this study. The first limitation is due to the somewhat subjective approach to creating the custom game levels included in Portal 2 Gf. Few sources are available for creating a simulated assessment tool, and I made several assumptions about how this process could be accomplished using effective game level design, setting and assessing game difficulty, and the decision to use Portal 2as an assessment instrument.

A second limitation was the limited progression of participants through game levels within the in-person sample. This lack of progression required an analysis that included the game's training levels rather than using only the planned evaluation levels. While the ability to learn the game was found to be related to Gf (and to a lesser extent, Gc), these relationships may change when performing a more strict evaluation. Despite this limit, the training levels themselves were continually increased in difficulty and the potential differences in scores as the result of their inclusion may be small.

Another limitation of the study is the lack of ethnic diversity within either sample. As one of the motivating factors in the development of this non-verbal measure of intelligence was to address adverse impact on racial and ethnic minorities, it was disappointing that recruitment efforts to address this imbalance were not successful and any effects associated with ethnicity could not be examined. The Wonderlic's small sample size also limited my comparisons and is likely to have affected observed correlations between both intelligence tests and the simulated game.

Future Research

Future research should build upon this study by recruiting a more ethnically- and sex-balanced sample to further investigate potential changes to subgroup scores attributable to the simulated assessment. Additional variables should also be examined to determine their potential impact on simulation performance. For example, measures of personality may be used as a predictor for test motivation, which may be a potentially meaningful factor when considering both test and simulation performance (J. E. Hunter & Schmidt, 1996).

Future versions of this assessment or similar tools should ensure that training of game concepts is kept to a reasonable time frame. In this study, participants were allowed to attempt all game levels in sequence during the entire administration period of the test, resulting in many undergraduate participants only finishing the initial training. In the context of this study, training levels could be combined, the number of puzzle elements could be reduced, or the assessment could focus entirely on the ability of participants to learn the puzzle elements. There is a wide array of potential modifications and I believe this to be one of the strongest capacities of simulated assessments.

Although I developed the Portal 2 Gf simulation to specifically tap into Gf, the effects of adverse impact from GMA measures may occur in other simulations and should be assessed to prevent potential adverse impact attributable to the use of simulations as assessment and training devices.

Conclusion

This study demonstrated relationships between tests of nonverbal (Gf), verbal (Gc), and *g* with performance data gathered from a modified commercial computer simulated game. Portal 2 Gf game performance was associated with tests of Gf, weakly with Gc, and not significantly related to *g*, demonstrating that Gf abilities can be approximated with simulated testing methods. The Portal 2 Gf simulation provided an opportunity to capture Gf abilities by requiring participants to express both inductive and deductive reasoning to complete game puzzles. Participants' prior gameplay frequency and self-reported skill (experience) was not found to affect the relationship between game performance and intelligence with any type of assessment in the study, indicating that Gf scores can be approximated with a simulated game regardless of a participant's game experience. These results suggest that computer simulated games may be a fruitful avenue for psychological testing, and a particularly useful one for employee selection needs, once a more extensive research base is established.

References

Batey, M., Furnham, A., & Safiullina, X. (2010). Intelligence, general knowledge and personality as predictors of creativity. *Learning and Individual Differences*, *20*(5), 532–535. doi:10.1016/j.lindif.2010.04.008

Beier, M. E., & Oswald, F. L. (2012). Is cognitive ability a liability? A critique and future research agenda on skilled performance. *Journal of Experimental Psychology. Applied*, *18*(4), 331–45. doi:10.1037/a0030869

Bell, B. S., & Kozlowski, S. W. J. (2008). Active learning: effects of core training design elements on self-regulatory processes, learning, and adaptability. *The Journal of Applied Psychology*, *93*(2), 296–316. doi:10.1037/0021-9010.93.2.296

Boyce, A. S., Corbet, C. E., & Adler, S. (2013). IGNITE Lightning Round: Simulations & Personnel Selection. In M. Fetzer (Ed.), *SIOP 2013* (pp. 1–26). Houston, TX.

Brody, N. (2007). Barriers to understanding racial differences in intelligence: Commentary on Hunt and Carlson (2007). *Perspectives on Psychological Science*, *2*(2), 214–215. doi:10.1111/j.1745-6916.2007.00038.x

Carretta, T. R. (2011). Pilot Candidate Selection Method. *Aviation Psychology and Applied Human Factors*, *1*(1), 3–8. doi:10.1027/2192-0923/a00002

Carroll, J. (1997). The three-stratum theory of cognitive abilities. In *Contemporary Intellectual Assessment: Theories, Tests, and Issues* (3rd ed., pp. 122–130). New York: Guilford Press.

Chan, D. (1997). Racial subgroup differences in predictive validity perceptions on personality and cognitive ability tests. *The Journal of Applied Psychology*, *82*(2), 311–20.

Chan, D., & Schmitt, N. (2004). An Agenda for Future Research on Applicant Reactions to Selection Procedures : A Construct-Oriented Approach. *International Journal of Selection and Assessment*, *12*(June), 9–23.

Dodrill, C. B., & Warner, M. H. (1988). Further studies of the Wonderlic Personnel Test as a brief measure of intelligence. *Journal of Consulting and Clinical Psychology*, *56*(1), 145–147. doi:10.1037//0022-006X.56.1.145

Duckworth, A., Quinn, P., Lynam, D., Loeber, R., & Stouthamer-Loeber, M. (2011). Role of test motivation in intelligence testing. *PNAS*, *108*(19), 7716–7720. doi:10.1073/pnas.1018601108

Engle, R., Tuholski, S., Laughlin, J., & Conway, A. (1999). Working memory, short-term memory, and general fluid intelligence: a latent-variable approach. *Journal of Experimental Psychology: General*, *128*(3), 309–331.

Fetzer, M., & Tuzinski, K. (Eds.). (2013). *Simulations for Personnel Selection* (pp. 1–268). New York: Springer. doi:10.1007/978-1-4614-7681-8

Gordon, H. W., & Leighty, R. (1988). Importance of specialized cognitive function in the selection of military pilots. *Journal of Applied Psychology*, *73*(1), 38–45. doi:10.1037//0021-9010.73.1.38

Gottfredson, L. S. (2000). Skills gaps, not tests, make racial proportionality impossible. *Psychology, Public Policy, and Law*, *6*(1), 129–143. doi:10.1037//1076-8971.6.1.129

Gutenberg, R. L., Arvey, R. D., Osburn, H. G., & Jeanneret, P. R. (1983). Moderating effects of decision-making/information-processing job dimensions on test validities. *Journal of Applied Psychology*, *68*(4), 602–608. doi:10.1037//0021-9010.68.4.602

Hakstian, a. R., Scratchley, L. S., MacLeod, A. a., Tweed, R. G., & Siddarth, S. (1997). Selection of telemarketing employees by standardized assessment procedures. *Psychology and Marketing*, *14*(7), 703–726. doi:10.1002/(SICI)1520-6793(199710)14:7<703::AID-MAR5>3.3.CO;2-9

Horn, J. L. (1968). Organization of abilities and the development of intelligence. *Psychological Review*, *75*(3), 242–59.

Horn, J. L., & Cattell, R. B. (1966). Refinement and test of the theory of fluid and crystallized general intelligences. *Journal of Educational Psychology*, *57*(5), 253–270.

Hughes, M. G., Day, E. A., Wang, X., Schuelke, M. J., Arsenault, M. L., Harkrider, L. N., & Cooper, O. D. (2013). Learner-controlled practice difficulty in the training of a complex task: cognitive and motivational mechanisms. *The Journal of Applied Psychology*, *98*(1), 80–98. doi:10.1037/a0029821

Hunter, J. E. (1986). Cognitive ability, cognitive aptitudes, job knowledge, and job performance. *Journal of Vocational Behavior*, *29*(1), 423–425.

Hunter, J. E., & Schmidt, F. L. (1996). Intelligence and job performance: Economic and social implications. *Psychology, Public Policy, and Law*, *2*(3-4), 447–472. doi:10.1037//1076-8971.2.3-4.447

Hunter, S. T., Cushenbery, L., & Friedrich, T. (2012). Hiring an innovative workforce: A necessary yet uniquely challenging endeavor. *Human Resource Management Review*, *22*(4), 303–322. doi:10.1016/j.hrmr.2012.01.001

Jentsch, F., & Bowers, C. (1998). Evidence for the validity of PC-based simulations in studying aircrew coordination. *The International Journal of Aviation …*, *8*(3), 243–260.

Johnson, W., & Bouchardjr, T. (2005). The structure of human intelligence: It is verbal, perceptual, and image rotation (VPR), not fluid and crystallized. *Intelligence*, *33*(4), 393–416. doi:10.1016/j.intell.2004.12.002

Jong, T. de. (1991). Learning and instruction with computer simulations. *Education and Computing*, *6*(1), 217–229.

Konig, C., Buhner, M., & Murling, G. (2005). Working memory, fluid intelligence, and attention are predictors of multitasking performance, but polychronicity and extraversion are not. *Human Performance*, *18*(3), 243–266.

Kroner, S., Plass, J., & Leutner, D. (2005). Intelligence assessment with computer simulations. *Intelligence*, *33*(4), 347–368. doi:10.1016/j.intell.2005.03.002

Lievens, F., Reeve, C. L., & Heggestad, E. D. (2007). An examination of psychometric bias due to retesting on cognitive ability tests in selection settings. *The Journal of Applied Psychology*, *92*(6), 1672–82. doi:10.1037/0021-9010.92.6.1672

Mayer, R. E. (2004). Should there be a three-strikes rule against pure discovery learning? The case for guided methods of instruction. *The American Psychologist*, *59*(1), 14–19. doi:10.1037/0003-066X.59.1.14

McGrew, K. (2009). CHC theory and the human cognitive abilities project: Standing on the shoulders of the giants of psychometric intelligence research. *Intelligence*, *37*(1), 1–10. doi:10.1016/j.intell.2008.08.004

McKay, P., & Doverspike, D. (2001). African-Americans' Test-taking Attitudes and Their Effect on Cognitive Ability Test Performance: Implications for Public Personnel Management Selection Practice. *Public Personnel Management*, *30*(1), 67–76.

Mead, A. D., & Drasgow, F. (1993). Equivalence of computerized and paper-and-pencil cognitive ability tests: A meta-analysis. *Psychological Bulletin*, *114*(3), 449–458. doi:10.1037//0033-2909.114.3.449

Mennecke, B., Hassall, L. M., & Triplett, J. (2008). The Mean Business of Second Life: Teaching Entrepreneurship, Technology and e-Commerce in Immersive Environments. *MERLOT Journal of Online Learning and Teaching*, *4*(3), 339–348.

Motowidlo, S. J., Dunnette, M. D., & Carter, G. W. (1990). An Alternative Selection Procedure : The Low-Fidelity Simulation. *Journal of Applied Psychology*, *75*(6), 640–647.

Naglieri, J., & Jensen, A. (1987). Comparison of Black-White differences on the WISC-R and the K-ABC: Spearman's hypothesis. *Intelligence*, *11*(1), 21–43. doi:10.1016/0160-2896(87)90024-9

Ortiz, G. (1994). Effectiveness of PC-based flight simulation. *The International Journal of Aviation Psychology*, *4*(3), 285–291.

Osato, R., & Sherry, P. (1993). Age and IQ effects on army enlisted male CPI profiles. *Military Psychology*, *5*(1), 41–61.

Ployhart, R. E., & Ehrhart, M. G. (2002). Modeling the Practical Effects of Applicant Reactions: Subgroup Differences in Test-Taking Motivation, Test Performance, and Selection Rates. *International Journal of Selection and Assessment*, *10*(4), 258–270.

Pynes, J., & Bernardin, H. (1989). Predictive validity of an entry-level police officer assessment center. *Journal of Applied Psychology*, *74*(5), 831–833.

Raven, J., Raven, J. C., & Court, J. H. (2003). *Manual for Raven's Progressive Matrices and Vocabulary Scales*. Harcourt: Harcourt Assessment.

Ree, M. J., & Earles, J. a. (1992). Intelligence Is the Best Predictor of Job Performance. *Current Directions in Psychological Science*, *1*(3), 86–89. doi:10.1111/1467-8721.ep10768746

Richardson, A. E., Powers, M. E., & Bousquet, L. G. (2011). Video game experience predicts virtual, but not real navigation performance. *Computers in Human Behavior*, *27*(1), 552–560. doi:10.1016/j.chb.2010.10.003

Richman-Hirsch, W. (2000). Examining the impact of administration medium on examinee perceptions and attitudes. *Journal of Applied Psychology*, *85*(6), 880–887. doi:10.1037//0021-9010.85.6.880

Roper, K., & Juneja, P. (2008). Distractions in the workplace revisited. *Journal of Facilities Management*, *6*(2), 91–109.

Roth, P. L., Bevier, C. a., Bobko, P., Switzer, F. S., & Tyler, P. (2001). Ethnic Group Differences in Cognitive Ability in Employment and Educational Settings: a Meta-Analysis. *Personnel Psychology*, *54*(2), 297–330. doi:10.1111/j.1744-6570.2001.tb00094.x

Roth, P. L., & Huffcutt, A. I. (2013). A Meta-Analysis of Interviews and Cognitive Ability. *Journal of Personnel Psychology*, *12*(4), 157–169. doi:10.1027/1866-5888/a000091

Rushton, J. P., & Jensen, A. R. (2005). Wanted: More race realism, less moralistic fallacy. *Psychology, Public Policy, and Law*, *11*(2), 328–336. doi:10.1037/1076-8971.11.2.328

Sager, C. E., Peterson, N. G., Oppler, S. H., Rosse, R. L., & Walker, C. B. (1997). An examination of five indexes of test battery performance: Analysis of the ECAT battery. *Military Psychology*, *9*(1), 97–120. doi:10.1207/s15327876mp0901_6

Schmidt, F. L. (2002). The Role of General Cognitive Ability and Job Performance: Why There Cannot Be a Debate. *Human Performance*, *15*(1-2), 187–210. doi:10.1080/08959285.2002.9668091

Schmidt, F. L., & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin*, *124*(2), 262–274. doi:10.1037//0033-2909.124.2.262

Schneider, W., & McGrew, K. (2012). The Cattell-Horn-Carroll model of intelligence. *Contemporary Intellectual Assessment: Theories, Tests, and Issues*, 99–144. Retrieved from http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:The+Cattell-Horn-Carroll+Model+of+Intelligence#1

Shipley, W. C., Gruber, C. P., Martin, T. A., & Klein, A. M. (2012). *Shipley-2 Manual*. Western Psychological Services.

Van Iddekinge, C. H., Roth, P. L., Raymark, P. H., & Odle-Dusseau, H. N. (2012). The criterion-related validity of integrity tests: an updated meta-analysis. *The Journal of Applied Psychology*, *97*(3), 499–530. doi:10.1037/a0021196

Waters, S. (2007). *General vs. specific ability tradeoffs in personnel selection: Consequences for validity and subgroup differences*. University of Minnesota.

APPENDIX A
TABLES

Table 1 Descriptive Statistics by Sample.

	Pilot			In-Person		
	n	*M*	*SD*	*n*	*M*	*SD*
Sex						
Male	87	-	-	31	-	-
Female	7	-	-	42	-	-
Ethnicity						
African American	0	-	-	9	-	-
Asian	2	-	-	2	-	-
Caucasian	85	-	-	53	-	-
Hispanic	3	-	-	5	-	-
Inter-Racial	2	-	-	4	-	-
Other	2	-	-	0	-	-
Age	94	24.7	4.3	73	18.6	1.3
Gameplay Experience	94	13.7	2.4	73	7.5	2.8
Portal 2 Performance Composite	94	.28	.38	73	-.91	.70
Wonderlic	27	28.6	2.9	-	-	-
Raven's Progressive Matrices	-	-	-	73	49.4	5.7
Shipley-2 Vocabulary	-	-	-	73	28.1	3.9
Shipley-2 Block Patterns	-	-	-	73	17.2	4.2

Blank spaces indicate assessments not administered to each sample.

Table 2 List of Game Objects in Order of Presentation.

Name of Object/Tool	**Passes Through Portals**	**Affected by Gravity**	**Powers Devices**	**Description***
Weighted Cube	Yes	Yes	Indirectly (Button)	A solid cube the player can carry that is used to weigh buttons down, stop lasers, or be pushed by the funnel.
Button	No	No (Attached to surface)	Yes	An immobile device, placed on any surface that powers another device when activated.
Laser	Yes	No	Yes	A red laser used to power devices, harms player if touched, can be stopped or redirected by a cube.
Reflective Cube	Yes	Yes	Indirectly (Reflects Laser)	Redirects laser if placed in path of laser beam.
Blue Goo	Yes	Yes	No	Surfaces touched with the goo become elastic, allowing objects to use floors and walls like trampolines. Can be redirected with the funnel.
Excursion Funnel	Yes	No	Indirectly (May push object onto button)	Pushes or pulls objects in one direction ignoring gravity, allowing objects to move straight up or across gaps without falling.

**note- the term 'objects' includes the player's character.*

Table 3 Correlation Matrix.

	1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*n*
1. Gameplay Experience	1	.79***	.21	.10	.10	.17	.78***	.57***	.58***	.87***	*167*
2. Portal 2 Performance Composite	.79***	1	.44***	.27*	.46***	.16	.95***	.89***	.90***	.88***	*167*
3. RPM	.21	.44***	1	.34**	.57***	*a*	.44***	.36**	.36**	.34**	*73*
4. Shipley-2 Vocabulary	.10	.27*	.34**	1	.34*	*a*	.26*	.15	.24*	.30*	*73*
5. Shipley-2 Block Patterns	.10	.46***	.57***	.34**	1	*a*	.43***	.34**	.41***	.37***	*73*
6. Wonderlic	.17	.16	*a*	*a*	*a*	1	.10	.09	.14	.27	*27*
7. Portal 2 - Time	.78***	.95***	.44***	.26*	.43***	.10	1	.79***	.82***	.80***	*167*
8. Portal 2 - Portals	.57***	.89***	.36**	.15	.34**	.09	.79***	1	.80***	.63***	*167*
9. Portal 2 - Steps	.58***	.90***	.36**	.24*	.41***	.14	.82***	.80***	1	.68***	*167*
10. Portal 2 - Number of Levels Completed	.87***	.88***	.34**	.30*	.37***	.27	.80***	.63***	.68**	1	*167*

Significance levels: *** $p < .001$, ** $p < .01$, * $p < .05$

a. Pilot Participants who completed the Wonderlic did not complete either the Shipley-2 or Raven's test.

Table 4 Multiple Linear Regressions Utilizing Game Performance and Game Experience as Predictors.

Measure	Model	Beta Coefficients				R^2	ΔR^2
		Portal 2 Composite		Game Experience			
		β	*b*(SE)	β	*b*(SE)		
RPM	1	.44	3.61(.87)			.20	.20***
	2	.45	3.63(1.0)	-.00	-.01(.25)	.20	.00
Shipley 2 - Block Patterns	1	.46	2.78(.64)			.21	.21***
	2	.53	3.20(.72)	-.15	-.23(18)	.23	.02
Shipley 2 - Vocabulary	1	.27	1.54(.64)			.08	.08*
	2	.29	1.64(.73)	-.04	-.06(18)	.08	.00
Wonderlic	1	.16	1.62(2.0)			.03	.03
	2	.09	.92(2.5)	.12	.26(.55)	.04	.01

*$p < .05$, ** $p < .01$, *** $p < .001$

Note: Model 1 Predictor: Game Performance Composite, Model 2 Predictors: Game Performance Composite and Game Experience

Table 5 Sex Comparison by Assessment within the In-person Sample.

	RPM	Shipley-2 Vocabulary	Shipley-2 Block Patterns
Men (*N* = 31)			
Portal 2 Composite	.49**	.33	.46*
Women (*N* = 42)			
Portal 2 Composite	.39**	.17	.45**

APPENDIX B

FIGURES

Figure 1 Comparison of Samples for Game Experience.

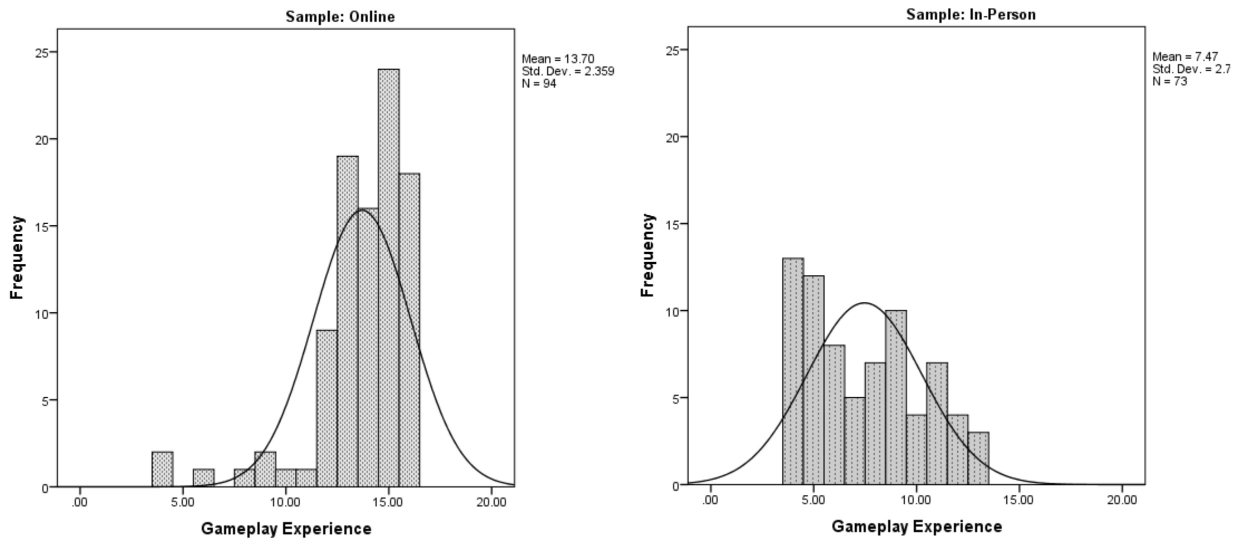
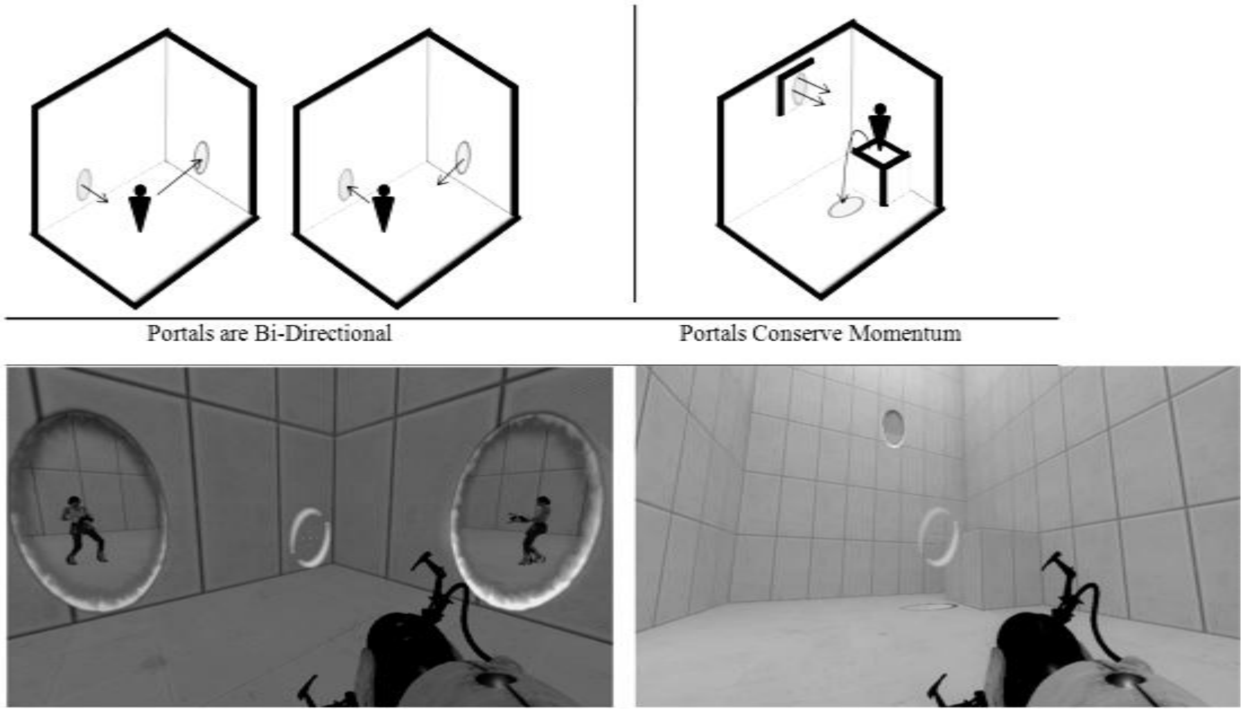
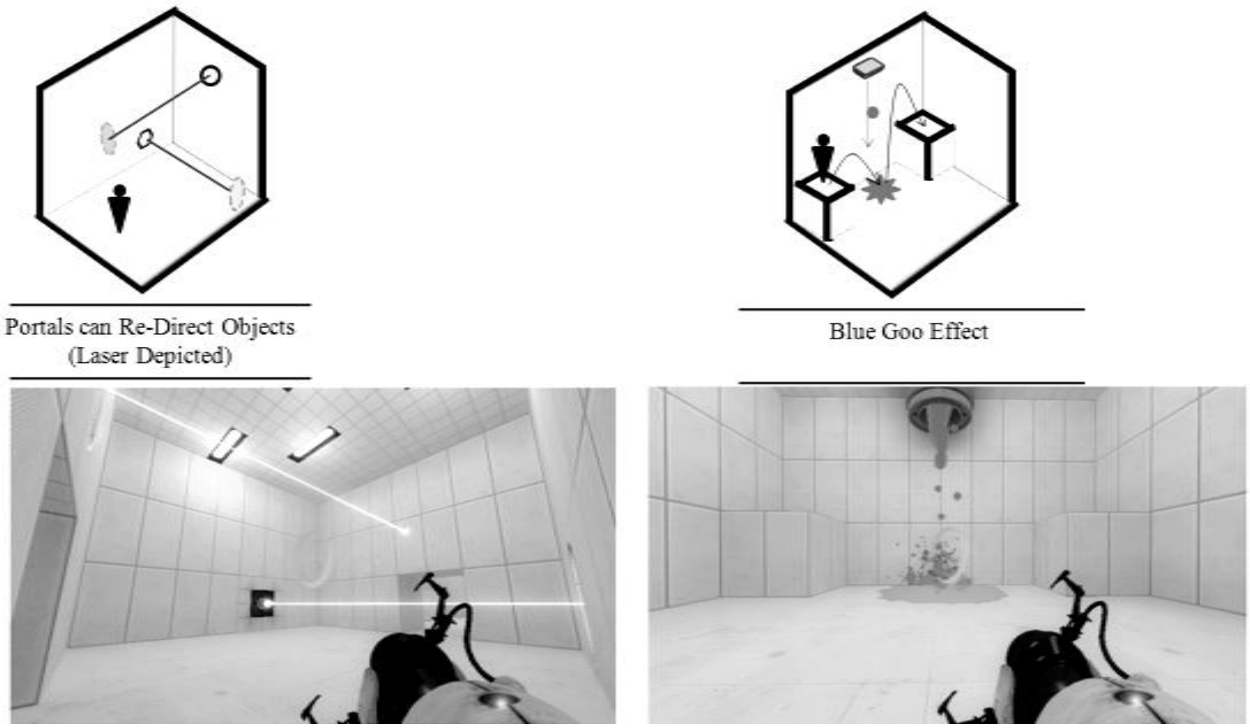


Figure 2 Portal 2 Basic Game Concepts - Portals.



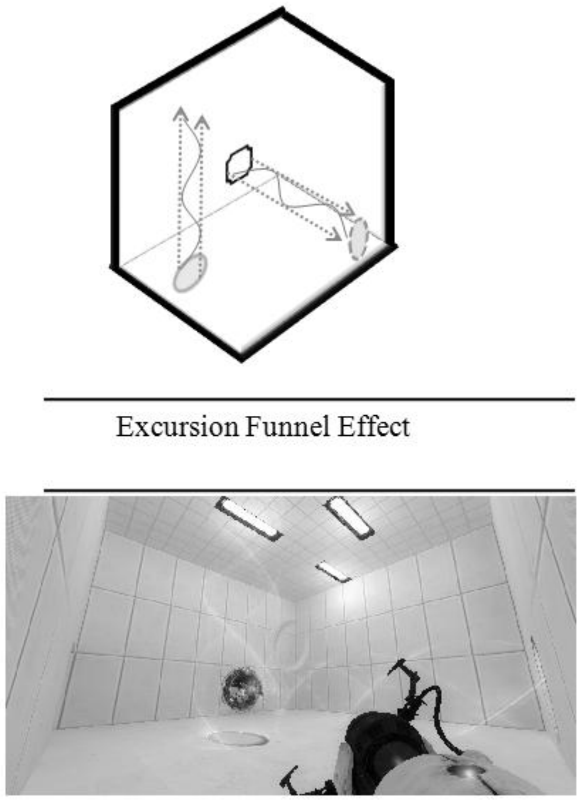
Note: These images represent the conceptual properties of portals within Portal 2. Above and Left: Conceptual depiction of a room with a player entering through either portal and exiting through the other. Below and Left: In-game view of this situation. Note that each portal shows the view from its exit, each depicting the player. Above and Right: Conceptual depiction of the use of inertia with a portal to transfer vertical velocity to a horizontal plane. Below and Right: In-game view of these placements.

Figure 3 Portal 2 Basic Game Concepts - Re-Directed Laser and Blue Goo.



Note: These images represent the Laser and Blue Goo puzzle elements within Portal 2. Above and Left: Conceptual depiction of a laser being transferred through a pair of portals by placing a pair of portals on the walls of the room. Below and Left: In-game view of this puzzle element. Above and Right: Conceptual depiction of using Blue Goo coating the floor to bounce across a gap. Below and Right: In-game view of this puzzle element.

Figure 4 Portal 2 Basic Game Concepts - Re-Directed Excursion Funnel.



Note: This image represents the Excursion Funnel puzzle element within Portal 2. Above: Conceptual depiction of a funnel being redirected from hitting a wall to hitting the ceiling by placing a pair of portals on the wall and floor. Below: In-game view of this puzzle element.

Figure 5 Projected Image of Portal 2 Controls.

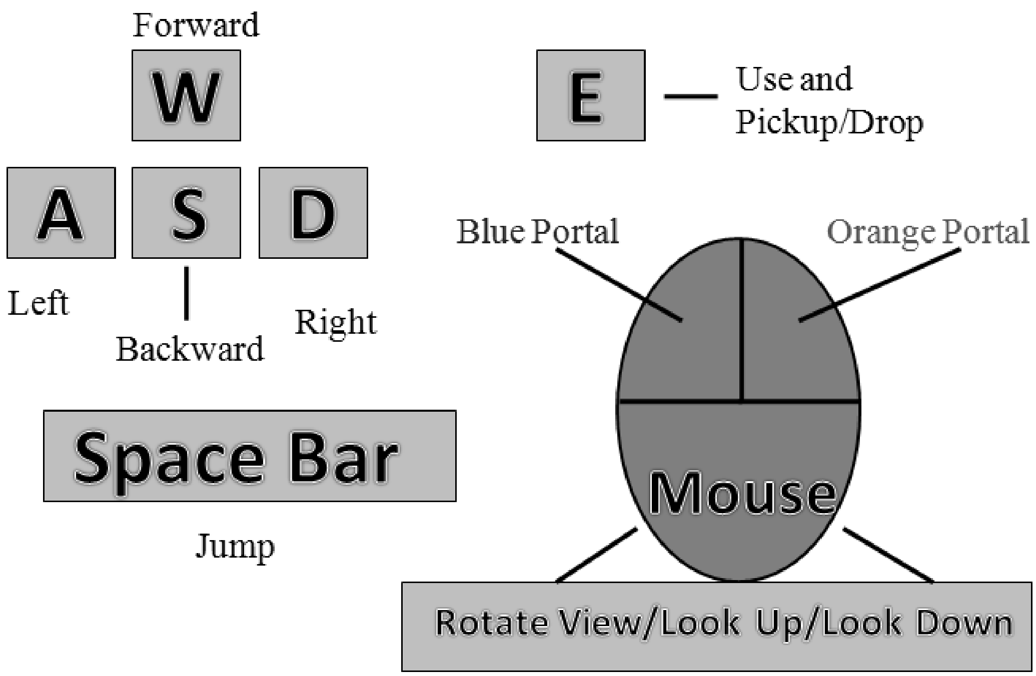


Figure 6 Number of Portal 2 Levels Completed by Sample.

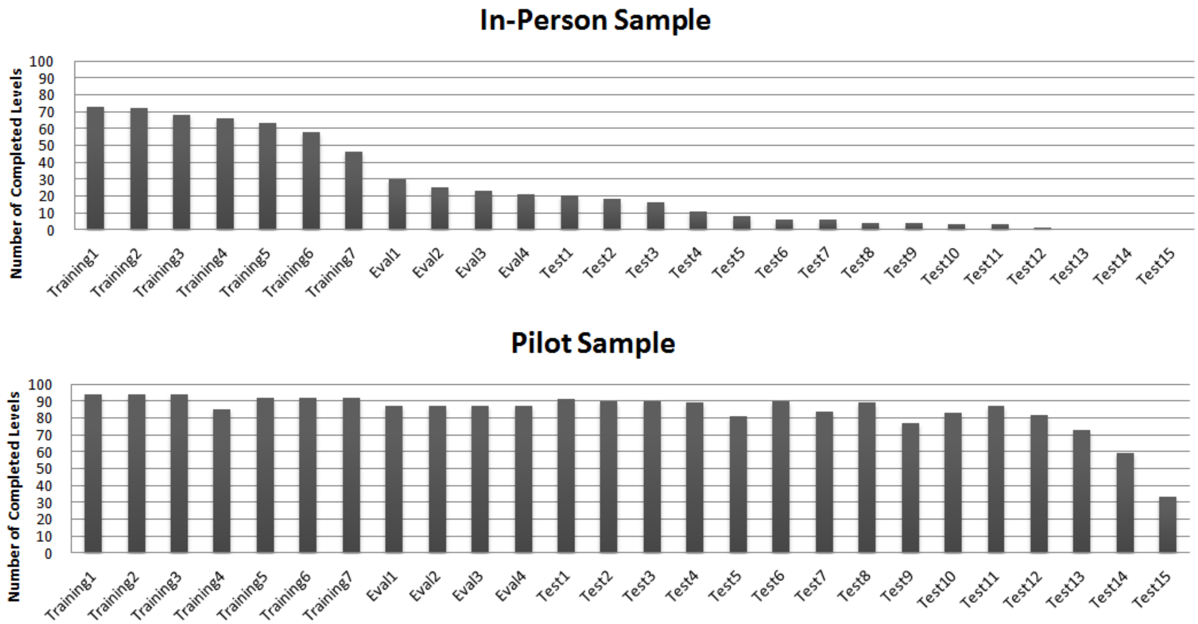
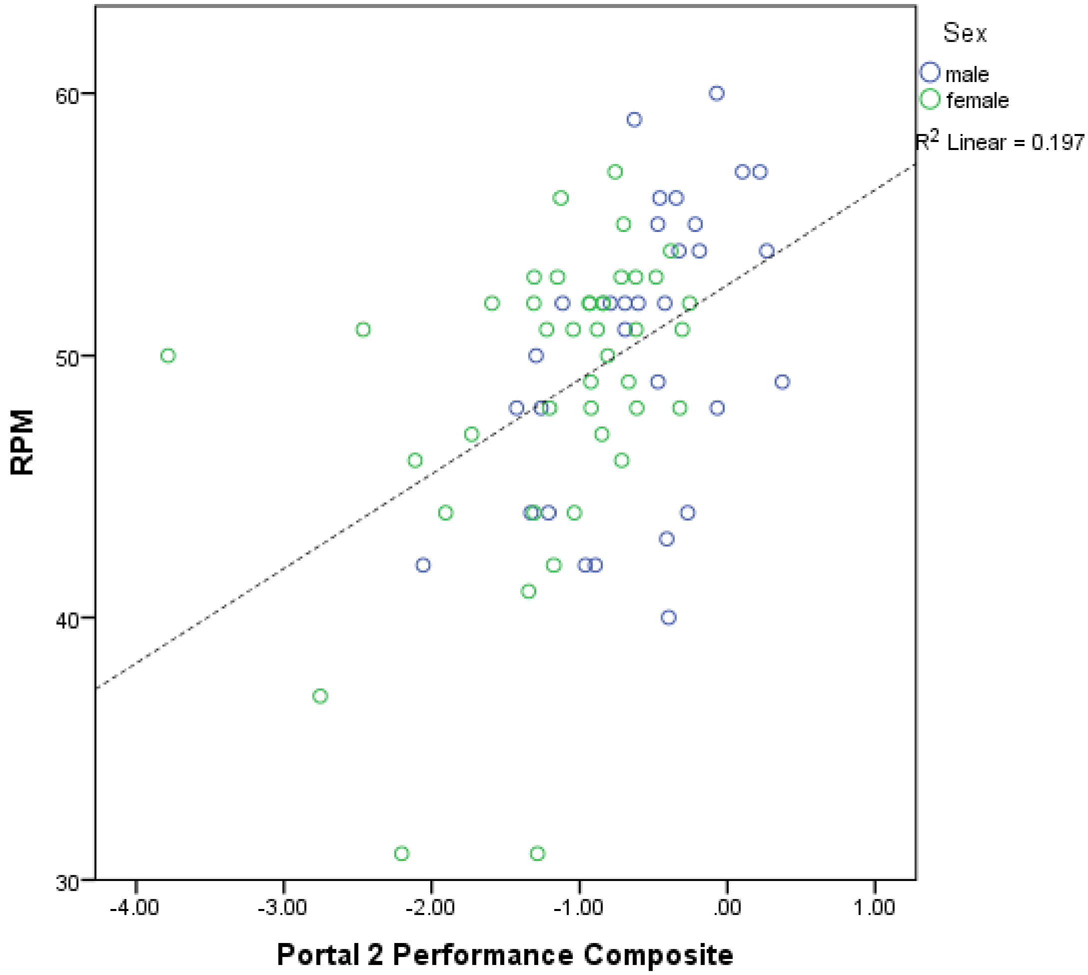


Figure 7 RPM and Time Performance by Gender.



APPENDIX C

IRB

THE UNIVERSITY of TENNESSEE CHATTANOOGA

Institutional Review Board
Dept. 4915
615 McCallie Avenue
Chattanooga, TN 37403-2598
Phone: (423) 425-5867
Fax: (423) 425-4052
instrb@utc.edu
http://www.utc.edu/irb

MEMORANDUM

IRB # 13-084

TO: Charles Buford
Dr. Christopher Cunningham
Dr. Michael Biderman
Dr. Joe Dumas
Dr. Brian O'Leary

FROM: Lindsay Pardue, Director of Research Integrity
Dr. Bart Weathington, IRB Committee Chair

DATE: June 5, 2013

SUBJECT: IRB# 13-084: Cognitive Ability and Advanced Simulation

The Institutional Review Board has reviewed and conditionally approved your application and assigned you the IRB number listed above. The following modification(s) need to be made to your research protocol before research may begin:

- Please include a statement on your consent form that addresses participant's reasonable expectation of confidentiality.

Please return a copy of the revised document to the IRB at instrb@utc.edu before beginning your research.

You must include the following approval statement on research materials seen by participants and used in research reports:

The Institutional Review Board of the University of Tennessee at Chattanooga (FWA00004149) has approved this research project #13-084

Please remember that you must complete a Certification for Changes, Annual Review, or Project Termination/Completion Form when the project is completed or provide an annual report if the project takes over one year to complete. The IRB Committee will make every effort to remind you prior to your anniversary date; however, it is your responsibility to ensure that this additional step is satisfied.

Please remember to contact the IRB Committee immediately and submit a new project proposal for review if significant changes occur in your research design or in any instruments used in conducting the study. You should also contact the IRB Committee immediately if you encounter any adverse effects during your project that pose a risk to your subjects.

For any additional information, please consult our web page http://www.utc.edu/irb or email instrb@utc.edu

Best wishes for a successful research project.

APPENDIX D

CONSENT FORM

CAAS
Consent Form

Date:__________________ Participant Number:________________

Dear University of Tennessee at Chattanooga Student Participant,
I am a Graduate Student under the direction of Dr. Brian O'Leary in the Department of Psychology at the University of Tennessee at Chattanooga. I am conducting a research study to determine the relationship between tests of ability and performance within a simulation. By participating in this experiment for any duration, or by forwarding your contact information to csbuford@gmail.com, you are automatically entered into a random drawing for a $100 Visa Gift Card.
I am requesting your participation for the completion of 1 session of data collection that is 3 hours long. During your the first half of the session you will be requested to complete 2 tests of ability, with a total allotted time of 1 hour and 30 minutes to complete both paper and pencil tests. A brief break will take place after these tests. During your second session you will be asked to complete a computer simulation for a total of 1 hour and 30 minutes on a computer. Your participation in this study is voluntary. If you choose to withdraw from the study at any time, there will be no penalty to any course credit or your chances of winning the Visa Gift Card drawing. The results of this study may be published, but your name and any individually identifiable information will not be used as a result of completing any part of this experiment. Your responses will be confidential as far as is possible under the law and access to response data will be limited to the research group.
Due to the nature of the simulation used in this study, if you have **any photosensitive medical condition** (e.g. **Epilepsy**) **or report motion sickness** with first-person simulations, you will not be allowed to participate due to medical concerns. If at any time you experience these symptoms, please notify the experimental administrator immediately. If you are under 18 years of age you will also not be allowed to participate.
If you have any questions concerning the research study, please contact Charles Buford at csbuford@gmail.com , or Dr. Brian O'Leary at Brian-O'leary@utc.edu or at (423) 425-4283. This research has been approved by the UTC Institutional Review Board (IRB). If you have any questions concerning the UTC IRB policies or procedures or your rights as a human subject, please contact Dr. Bart Weathington, IRB Committee Chair, at (423) 425-4289 or email instrb@utc.edu.
Signature and submission of this document will be considered your consent to participate.

Thank you!

Your Printed Name: ____

Your Signature:____

The Institutional Review Board of the University of Tennessee at Chattanooga (FWA00004149) has approved this research project #13-084

APPENDIX E

DEMOGRAPHIC SURVEY

Information Form

Please complete the following questions to the best of your ability.
Your responses will remain confidential.

Full Name (First and Last)

Your Participant Number (Found at the top of your consent form)

What is your computer number (located on the top of the desk to the left)

Preferred method of contact (Phone Number or Email)

What is your current age?

What is your biological sex?

- Male
- Female

Your Ethnicity?

- African American
- Asian
- Caucasian
- Hispanic
- Inter-racial
- Other (Not Listed)

How often do you play electronic games (video games) on any game platform on average?

- I do not normally play video games
- About a couple times per year
- About a couple times per month
- About once a week
- About once a day
- At least once per day

How skilled would you rate yourself at electronic games (video games) ?

- Unskilled
- Average
- Skilled

How skilled would you rate yourself at electronic puzzle games (video games) ?

- Unskilled
- Average
- Skilled

How skilled would you rate yourself at Portal or Portal 2?

- I have not played Portal or Portal 2
- Unskilled
- Average
- Skilled

Please only hit the Send button once.

Send Clear

VITA

Charles Buford was born in Huntsville, AL to Belinda and Keith Buford and is the older brother to Cole Buford. He attended the University of Alabama in Huntsville where he became interested in psychology. Working closely with Dr. Eric Seemann, he developed an interest in psychological assessment tools and methods. He completed his Bachelor of Arts in Psychology during May 2012 before beginning his studies at the University of Tennessee at Chattanooga's Master of Science in I-O Psychology. While working with Dr. Brian O'Leary, he developed an appreciation for the applied needs of scientific research. While at UTC Charles was also able to serve as the president of the university's graduate students. Charles hopes to graduate with this degree and begin a career in industry.