THE APPLICATION OF COGNITIVE DIAGNOSTIC APPROACHES VIA NEURAL NETWORK ANALYSIS OF SERIOUS EDUCATIONAL GAMES

by

Richard L. Lamb

A Dissertation
Submitted to the
Graduate Faculty
of
George Mason University
in Partial Fulfillment of
The Requirements for the Degree
of
Doctor of Philosophy
Education

Committee:

_____ Chair

_____

_____

_____ Program Director

_____ Dean, College of Education
and Human Development

Date: _____ Summer Semester 2013
George Mason University
Fairfax, VA

The Application of Cognitive Diagnostic Approaches via Neural Network Analysis of Serious Educational Games

A Dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy Education at George Mason University

by

Richard L. Lamb
Doctor of Philosophy
George Mason University, 2013

Director: Leonard A. Annetta, Associate Professor
College of Education and Human Development

Summer Semester 2013
George Mason University
Fairfax, VA

DEDICATION

I dedicate my dissertation work to Rebekah, Trinity my daughter, Kyler my son; I love you all. I would like to express a feeling of love and gratitude to my mother, Joyce, my grandmother Myrna and my grandfather, Theodore, without their loving guidance throughout my life I would not have come this far. I also would like to thank my brother Robert, Uncle John and Aunt Kathy for their encouragement when I needed it most.

I also dedicate this dissertation to the rest of my family and friends who have supported me throughout this process. To Katherine, Mike, and Maggie who have provided friendship, feedback, and levity throughout this process and others who have pushed and pulled me to finish when I needed it.

ACKNOWLEDGEMENTS

I would like to thank my committee for their feedback. In particular my mentor Leonard Annetta who has always, without fail, guided me through my Master and Doctorate pursuits and pushed me to accomplish when I did not know I could. I would like to thank Dimiter Dimitrov for his guidance in measurement and research methods and Anastasia Kitsantas for her understanding and suggestions in Educational Psychology as I worked toward my goals. I would also like to acknowledge Jose Sandoval at Duke University and Elliot Inman at SAS for igniting my interest in modeling and quantitative analysis.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF EQUATIONS

LIST OF ABBREVIATIONS AND SYMBOLS

Activation Function ..... φ
Attribute ..... Ai
Attribute Mastery Model Fit ..... HCl_i
Attribute Probability ..... P_{Ai}
Coefficient of Agreement ..... d
Coefficient of Determination ..... R^2
Delta (Change) ..... Δ
Eta (Normalization Factor) ..... η
Gradient Decent ..... D
Individual Item ..... Xi_j
Item Category Count I
Item Completion Probability ..... Pi
Item Response Function ..... Π
Lambda ..... Λ
Neural Network Propagation Weight ..... Wi
Neural Network ..... N_c
Output Neuron Weight ..... Y
Sigma (Summation) ..... Σ
Subject Ability ..... Θ

ABSTRACT

THE APPLICATION OF COGNITIVE DIAGNOSTIC APPROACHES VIA NEURAL NETWORK ANALYSIS OF SERIOUS EDUCATIONAL GAMES

Richard L. Lamb, Ph.D.

George Mason University, 2013

Dissertation Director: Dr. Leonard A. Annetta

Serious Educational Games (SEGs) have been a topic of increased popularity within the educational realm since the early millennia. SEGs are generalized form of Serious Games to mean games for purposes other than entertainment but, that also specifically include training, educational purpose and pedagogy within their design. This rise in popularity (for SEGs) has occurred at a time when school systems have increased the type, number, and presentations of student achievement tests for decision-making purposes. These tests often task the form of end of course (year) tests and periodic benchmark testing. As the use of these tests, has increased policymakers have suggested their use as a measure for teacher accountability. The change in testing resulted from a push by school districts and policy makers at various component levels for a data-driven decision-making (D3M) approach. With the data-driven decision making approaches by school districts, there has been an increased focus on the measurement and assessment of student content knowledge with little focus on the contributing factors and cognitive attributes within

learning that cross multiple-content areas. One-way to increase the focus on these aspects of learning (factors and attributes) that are additional to content learning is through assessments based in cognitive diagnostics. Cognitive diagnostics are a family of methodological approaches in which tasks tie to specific cognitive attributes for analytical purposes. This study explores data derived from computer data logging (n=158,000) in an observational design, using traditional statistical techniques such as clustering (exploratory and confirmatory), item response theory and through data mining techniques such as artificial neural network analysis. From these analyses, a model of student learning emerges illustrating student thinking and learning while engaged in SEG Design. This study seeks to use cognitive diagnostic type approaches to measure student learning while designing science task based SEGs. In addition, the study suggests that it may be possible to use SEGs to provide a means to administer cognitive diagnostic based assessments in real time. Results of this study suggest the confirmation of four families (factors) of traits illustrating a simple factor loading structure. Item response theory (IRT) results illustrate a 2-parameter logistic model (2PLM) fit allowing for parameterization using the IRT-True Score Method (χ^2=1.70, *df*=1, *p*=0.19). Finally, fit statistics for the artificial neural network suggest the developed model adequately fits the current data set and provides a means to explore cognitive attributes and their effect on task outcomes. This study has developed a justification for combining and developing two distinct areas of research related to student learning. The first is the use of cognitive diagnostic approaches to assess student learning as it relates to the cognitive attributes used during science processing. The second area is an examination and modeling of the relationship

between attributes as propagated in an artificial neural network. Results of the study provide for an ANN model of student cognition while designing science based SEGs (r^2=0.73, RMSE= 0.21) at a convergence of 1000 training iterations. The literature presented in this dissertation work integrates work from multiple field areas. Fields represented in this work range from science education, educational psychology, measurement, and computational psychology.

CHAPTER ONE

Introduction

One goal of science educators is to assist secondary school students (9-12) in achieving increased levels of understanding in their learning content areas, specifically the natural sciences. One potential way to improve this understanding is with the use of Serious Educational Games (SEG). SEG computer games and their closely related brethren, virtual laboratory simulations, are of immediate interest to the science education community. SEGs are games designed for educational or training purposes with specific pedagogical approaches. The inclusion of pedagogical approaches specifically differentiates SEGs from other forms of computer-based learning. A second way that SEGs differ from their counterparts is by the immersive nature of their environments. Within the SEG, the virtual learning environment mimics the actual environments in which the subject would conduct their tasks as closely as possible. The mimicry of actual environments along with the interactivity (open-ended) and complexity of the environment and problems is what adds to the authenticity of the tasks (Annetta, 2010). The task authenticity provides an ideal means for SEGs to present virtual tasks to subjects for diagnostic purposes. In addition to the task authenticity, SEGs provide an additional advantage of creating continuous, large data streams through data logging for analysis.

Given the digital nature of the data stream, SEG data readily lends itself to effective Bayesian data mining techniques such as Artificial Neural Network Analysis.

Video Games in Education

One goal of science educators is to assist secondary school students (9-12) in achieving increased levels of understanding in their learning content areas, specifically the natural sciences. One potential way to improve this understanding is with the use of Serious Educational Games (SEG). SEG computer games and their closely related brethren, virtual laboratory simulations, are of immediate interest to the science education community. SEGs are games designed for educational or training purposes with specific pedagogical approaches. The inclusion of pedagogical approaches specifically differentiates SEGs from other forms of computer-based learning. A second way that SEGs differ from their counterparts is by the immersive nature of their environments. Within the SEG, the virtual learning environment mimics the actual environments in which the subject would conduct their tasks as closely as possible. The mimicry of actual environments along with the interactivity (open-ended) and complexity of the environment and problems is what adds to the authenticity of the tasks (Annetta, 2010). The task authenticity provides an ideal means for SEGs to present virtual tasks to subjects for diagnostic purposes. In addition to the task authenticity, SEGs provide an additional advantage of creating continuous, large data streams through data logging for analysis. Given the digital nature of the data stream, SEG data readily lends itself to effective Bayesian data mining techniques such as Artificial Neural Network Analysis.

There is considerable research, which, suggests that when properly designed with underlying science concepts and by extension, Science, Technology, Engineering, and Mathematics (STEM) principles, learning environments improve student skill levels in cognitive attributes such as practical reasoning, complex problem solving, transfer of learning, inductive reasoning, and the use of mapping in multidimensional space (Abell & Lederman, 2007; Spector & Changmin, 2012). In addition to improving student learning using videogames, there is considerable interest within the educational, government and business community in developing game formats that are capable of assessing students' science and STEM learning outcomes (Wall, 2011). Assessment of student learning is a multi-billion dollar industry with many stakeholders within and outside of, the educational community (Flaitz, 2011). Assessment is also one of the most contentious issues within the current educational environment with educators and policymakers often lining up on opposite sides as to the role of assessment within the school system (Messick, 1985; Odena, 2010). Educators suggest that the assessments are not meaningful measures of student learning and do not account for key student gains. On the other side of the debate about assessments role, policy makers demand accountability of the educator as the primary function of the test. This disconnect between educators and policymakers provides a stimulus for business and government to seek more appropriate and authentic measures of student learning in order to drive decision making processes for curriculum and learning (Demarest, 2010; Young, 2011). The call by business, government, educators, and policymakers for more authentic and realistic assessment has driven much of the innovation in assessment in recent years (Hanson & Mohn, 2011;

Hall, 2012). On this point, Educational Testing Service (ETS) presented the ETS Assessment Games Challenge as a means to challenge researchers and educators to develop assessments tasks in the form of a game. ETS presents awards on a yearly basis around assessed learning progressions (ETS, 2012). ETS is not the only entity to call for researchers to help bridge the gap in assessment, the National Science Foundation (NSF), Microsoft, Intel, the National Science Teacher Association and the National Education Association have all committed resources to address the disconnect in assessment between educators and policymakers. Each of these organizations sees educational gaming (SEGs) as a means to accomplish the repair of this disconnect and meaningfully engage students in the assessment process.

Background, Previous Work and Justification

There is very little argument that high stakes content assessments affect the science curriculum tremendously through its choice of question and topic (Penuel, Fishman, Gallagher, Korbak, & Lopez-Prado, 2009). The power of the "test" rests with the tests ability to shape, stretch, and remove portions of the science curriculum. The test's effect on the curriculum has been a subject of discussion since the inception of testing. The origins of formal testing are thought to begin with the Chinese Imperial Examination system designed to select candidates to serve as administrators as early as 141 BCE during the Han Dynasty (Elman, 2002). Impacts on the Chinese curriculum resulting from these examinations was well established by 105 BCE with the mandate from Emperor Wu of Han that all local officials take part in examinations designed to show aptitude through an understanding of the Confucian Classics. To say high stakes

testing has been a part of education for an extended period is an understatement. From this beginning, the concept of the standardized test to measure content extended through the Chinese sphere of influence and eventually arose in the West based upon the Chinese Imperial Examination. Viewing documents from the late 15th and 16th century, there are hints to the use of testing within the Western hemisphere with the development of entrance examinations for university admissions and the origins of the doctoral dissertation and thesis arising during this time (Clegg, 1979). Testing in this capacity remained relatively unchanged through the 17th and 18th century into the early 19th century. At this point it is necessary to focus away from a global-historical view of testing due to the rise of the dominance of empirical study such as mathematics, science etc. Specifically, it is necessary to focuses on the rise of science education as it development intertwines with the development of the testing movement within the United States.

Prior to reform in the early 19th century, science education within the United States was often a subjective scattering of practical topics such as navigation, agriculture, and surveying. The beginning of the standardization of the science curriculum occurred when Harvard University required the completion of high school courses in Physics and Chemistry for admission. Shortly thereafter, many colleges and universities followed Harvard's lead. In essences, the high school science courses developed into the premier admissions requirements for the universities. In 1892, the Committee of Ten, a group of educators designated to appoint subject matter experts and make recommendations for curriculum and college admission requirements, appointed three subject matter experts in

the sciences. The science content areas specifically represented were Physics, Astronomy, and Chemistry. The members of the Committee of Ten submitted several recommendations via the *Cardinal Principals of Secondary Education*, to the National Education Association for implementation (NEA, 1918). Chief among the recommendations were (a) elementary science should focus on natural phenomena, (b) secondary science should focus on laboratory work, and finally (c) science is a means to develop students for college. This began to set science apart from other content areas and set the stage for science to act as an assessment for college admissions. A second set of reforms in 1920 called the *Reorganization of Science in Secondary Schools* added to the role of science specifically the reforms called to develop individuals for effectiveness in science (NEA, 1920).

Over time, the science programs found their place not just within the university setting but also within the K-12 setting and science-teacher education programs. Acceptance of science education as a discipline came about with the inclusion of science education methods courses within college and university departments of education. As science education progressed, it began to change, after 1957 with the insistence by leaders within the United States that the United States match the Soviet Union scientifically and technologically after the launch of Sputnik 1 and 2 in 1957. As 1960 approached, the science education community shifted its focus from individual preparation to a strategic view of science's role within the country with the National Defense Education Act (1960) and the Guidelines for Secondary Science and Mathematics (1961). Radical new technologies and concerns for national security drove

the science educational reforms of the 1960s and early 1970s. However, the educational community provided an inadequate response to these developing scientific aspects of society and this initiated the shortage of technically trained personnel the educational community is grappling with today. As the acute need for STEM trained professionals has grown, the United States has declined in its position economically and educationally. This decline is exemplified in the Trends in Mathematics and Science Survey (TIMSS) from 1995 through 2011. As seen in an analysis of United States performance trends, the United States made the top 10 rankings in science only in the years, 2007 (9^{th}), and, 2011 (10^{th}) (United States Department of Education, 2013).

As part of the growing need for adequately trained personnel several science education stakeholders' organizations organized reforms. The Department of Education in 1983 released *The Nation at Risk: An Imperative for Educational Reform* outlining the need for radical educational change across all content areas. In an attempt to remedy the inadequate science education response of the 1970s (outlined in the *Nation at Risk* Report), the National Science Teacher Association (NSTA) constructed the science-technology-society (STS) curriculum. The goal of STS was to provide educators a means to develop scientifically minded individuals. With the position of the NSTA in mind, The American Association for the Advancement of Science (AAAS) implemented Project 2061 in 1985. Project 2061 was an attempt to identify the most important aspect of science education. Reformers at AAAS outlined many of the most important recommendations of Project 2061 in the 1989 publication, *Science for All Americans*. An additional inclusion within the Project 2061 reform that was not included in other reforms

was the inclusion of informal education as a contributing equal partner to science education.

During the development of the science education reforms of the middle and late 1980s, a parallel movement occurred; this movement was the rise of required national benchmarking. Congress established the National Assessment Governing Board to set policy for the National Assessment of Education Progress (NAEP) or more familiarly known as The Nation's Report Card. Following NAEP was Goals 2000: Education America Act. The Goals 2000 Act provided the initial foundations for standards based educational reform, and provided the framework for the development of the No Child Left Behind Act, 2001 (NCLB). It is at this point that assessment as envisioned by the Imperial Exams of China and throughout educational history and modern assessment depart from one another. Assessment moved from a means to understand student content learning to a means to hold educators, schools, and state public educational agencies accountable. Policy makers and educators have particularly focused on accountability in science education as a means to reform education and increase the number of technically proficient citizens. Common Core –developed by the National Governors Association Center for Best Practices- and the 21st Century Skills (Trilling & Fadel, 2009), and the Next Generation Science Standards (Achieve Inc., 2013) holds science as a corner stone of economic development and as a means to maintain the United States' current role in the world.

Within the current context of the testing and accountability, movement there seems to be a dichotomy between content learning measurement and stakeholder

accountability. The standards focus more on cognitive approaches, while assessments for accountability focus more on content. This dichotomy can be resolved though the development of new assessment techniques. The focus on accountability and assessment, coupled with the rise of inexpensive computing power, has increase the frequency and amount of data collected on students. Increased data collection has led to calls by those outside of education to mirror professions in which data provides a means to establish empirically based rational decisions, in other words, a Data Driven Decision Making (D3M) approach. The rise of D3M occurred in parallel with the rise of the accountably movement in education. D3M refers to the use of data to inform educational decisions. Specifically, No Child Left Behind (NCLB) mandates that the educational units specifically gather, aggregate, and report student-level data. NCLB also implicitly demands that school units initiate change in teaching practice, based upon changing accountability data, into actionable information. Many educators and educational units lack the specific training in analysis and data mining to make use of the vast data collected from assessments. Thus, there are vast data streams that are currently underutilized. Recently however, there have been movements by private organization and universities to develop a new field by combining analysis of data with pattern seeking analysis with continuous, large, data streams. Tentatively, this field has been named educational analytics or educational informatics. These areas are on the forefront of assessment and D3M. This dissertation is an attempt to provide a methodological approach centered on student cognition as a means to analyze and develop a sound D3M approach for science education.

Current assessment techniques both in education and in psychology rely heavily upon the theoretical foundations of Classical Test Theory (CTT) or alternatively, Item Response Theory (IRT). While each of these theoretical frameworks provides for a meaningful approach to test development and assessments, in general terms, other more recently developed techniques (beginning in the late 1980s) may prove useful. Chief among these techniques is that of Cognitive Diagnostics used as an assessment technique. Thus, we seek to separate and target the underlying aspects of learning through the targeting of student thinking. A somewhat analogous understanding of the relationship between content and cognitive attributes is, content is the phenotype (outward expression of learning) while cognition is the genotype (the underlying expression of learning). Using cognitive diagnostic approaches, coupled with neural network modeling, it may be possible to develop assessments using authentic task presentation with fewer items and outcomes for analysis across multiple curricular domains. These alternative assessment types would take the form of the design of SEGs using cognitive diagnostics.

Purpose, Research Questions and Hypothesis

The primary purpose of this study is to establish and suggest a methodological procedure for exploration of large data sets and analysis of the latent cognitive attributes associated with the design of SEG as a means to teach science concepts. The second goal of this study is to understand the proper data structures, which allows an artificial neural network (ANN) model to converge and provide a meaningful simulation of student learning relating to SEG design. The third purpose of this study is to uncover the associated cognitive attributes impacted while designing science based SEGs. Lastly, this

study seeks to develop and propose an exploratory model of the interaction between the cognitive attributes, factors, and game items. The research questions and associated hypotheses addressed within this study are:

Research Question 1 (RQ1): What are the underlying factors exhibited through the measurement of task items associated with student development of Serious Educational Games?

Hypothesis 1 (H1): Using appropriate clustering techniques, one can cluster Serious Education Game tasks to provide meaningful information for the development of a computational model. H_o $\Lambda_i=0$

Research Question 2 (RQ2): What are the cognitive attributes that underlie the design of Serious Educational Games?

Hypothesis 2 (H2): By using Serious Educational Game design, in the learning environment, it is possible to map the relationship between task items and latent cognitive attributes. H_o $P_{Ai}=0$

Research Question 3 (RQ3): What theoretical mathematical / statistical model develops using an Artificial Neural Network to describe the interaction of the items, factors, and attributes as student design Serious Educational Games?

Hypothesis 3 (H3): A computational-cognitive model to describing the underlying cognitive attributes activated while designing Serious Educational Games can be developed with valid predictive value. H_0 $R^2=0$

As researchers discover the latent attributes driving cognition, they can prescribe tasks to stimulate cognitive attributes associated with the science-based tasks. The key process for discovering the latent cognitive attributes is with cognitive diagnostic

approaches. This study presents literature from a series of papers examining how the design of games acts as an assessment tool and proposes the underlying cognitive attributes activated when engaging in tasks related to science game design and play. However, many of the current games used for educational purposes lack the depth and intricacy found in SEGs. Current work and literature completed by researchers thus far, has focused on the two variations of the second research question. The first is what factors affect science based game play, and what role does science based games play in student learning outcomes. For the completion of this study, the study tested each of the components related to SEG design. Each component has been tested using techniques outlined in the methods section and validated through a pilot study.

The remainder of the dissertation covers many topics to include computational modeling, measurement, and cognition. The Background section defines SEGs and makes the case for using artificial neural networks (ANN) and cognitive diagnostic approaches as a means to model and study student cognition. The Background also introduces and justifies the choice of SEGs in science education as a domain topic for experimentation. The Model section introduces neural networks for the application of cognitive diagnostics. The Methodologies section presents means for developing a picture of cognitive attributes and task completion. The Results section describes analytical outcomes and develops a proposed model for student cognition. The Future Work section describes how to continue the direction of this research to understand the conceptual basis of the observed results. In the final two sections, the study discusses the results of the cognitive diagnostic analysis and presentations an artificial neural network

model as a potential form of assessment. This form of assessment contrasts with other assessment approaches such as Item Response Theory and Classical Test Theory.

Definitions

The study contains several key words and concepts, which have specific contextual meanings and operationalized for the purposes of this study. The following section contains key definitions as used within the context of this study.

A priori: A method of relating to, or denoting reasoning or knowledge proceeding from theoretical deduction as opposed to, observation or experience (Demopoulos, 2003).

AAAS: The American Association for the Advancement of Science.

Artificial Neural Network: A mathematical model consisting of interconnected groups of neurons used for information processing and predicting outcomes (Oczkowski & Barreca, 1997).

Attribute Mastery Pattern (AMP): The pattern denoting the presence or absence of particular cognitive attributes related to a task. The AMP is similar to item response pattern s within the IRT framework, represented by the symbol P_A (Im & Yin, 2009).

Bartlett Test of Sphericity: A test statistics used to examine the hypothesis that the correlation matrix is an identity matrix meaning the variables are uncorrelated (Yang, 2005).

Bayesian Models: A framework for probabilistic inference providing a general approach to understanding problems of induction paralleled in the development within the human mind (Peifors, Tenenbaum, Griffiths, & Xu, 2011).

Classical Test Theory: A testing theory in which assumes that each observed score (X) contains a True score component (T) and an Error Component (E) (Lord, 1980).

Cluster Analysis: A statistical technique that naturally group data using response relationships (Borgen & Barnett, 1987).

Cognitive Architecture: A computer architecture involving multiple inference process within artificial neural network software, made to model the human brain using fuzzy logic calculations (Papageorgiou, 2011).

Cognitive Attributes: Psychological characteristics shown to be use during the thinking process. Attributes are normally distributes and relatively stable (Riding & Cheema, 1991).

Cognitive Matrix: See Q-Matrix

Confirmatory Factor Analysis: A form of factor analysis used in social science research to measure fit to a conceptual model (Fabrigar, Wegener, MacCallum, & Strahan, 1999).

Conjunctive Model: A cognitive diagnostic model, which assumes all attributes, must be present in some degree in order to complete a task (Medin & Shoben, 1998).

Construct Validity: The degree to which one can infer from operationalized definitions (Fraenkel, Wallen, & Hyun, 1993).

Critical Reasoning: A cognitive attribute exemplified by the purposeful use of information to, informed outcomes focused using logic, skills, and experience (Phillips & Bond, 2004).

Cronbach's Alpha: The average inter-correlation coefficient among items, this coefficient denotes levels of reliability within measures and assessments (Cortina, 1993).

Data Driven Decision Making (D3M): A process of curriculum and instructional development based upon the analysis of student level assessment data (Mandinach, Rivas, Light, Heinze, & Honey, 2006).

Digital Game-Based Learning: Learning games on a computer or online which are developed to teach content (Papastergiou, 2009).

Discrete Latent Attribute Model: A statistical model relating hidden variable values to observable variables (Meila & Heckerman, 2001).

Edutainment: Entertainment, computer games, films, or shows designed with educational aspects (Squire, 2003).

E-Learning: Learning occurring online or via the Internet using a computer (Sharma & Kitchens, 2004).

Exploratory Factor Analysis: A statistical technique used to identify clusters of inter-correlated variables (Goddard & Kirby, 1976).

Eigenvalues: The explained variance of the factor loadings in factor analysis (Tinsley & Tinsley, 1987). .

Fit Statistics: The set of statistics describing the measures the total deviation of the responses values to the expected values (Smith, 1991).

Frontolimbic: The portion of the brain responsible for attention and arousal (Fichtenholtz, Dean, Dillon, Yamasaki, McCarthy, & LaBar, 2004).

Frontostriatal: See Frontolimbic

Functional Magnetic Resonance Imaging (fMRI): A method of brain imaging designed to detect metabolic changes in brain function meant to represent activity within the brain (DeYoe, Bandettini, Neitz, Miller & Winans, 1994).

Gaussian Distribution: A normal distribution with a standard deviation of one and a total area under the curve equal to one this distribution is the basis for many inferential statistics (DeLong, DeLong, & Clarke-Pearson, 1998).

Generalized Linear Model: A generalization of the linear regression model used for modeling and prediction of relationships between variables (Nelder & Wedderburn, 1972).

GRADUATE: A National Science Foundation grant funded project examining the role and effects of Serious Educational Game design on high-risk students (Annetta, 2008).

Hebb's Synapse and Learning Rule: A model used to explain associative learning and the activation of neurons (Caporale & Dan, 2008).

Hidden Nodes: A Node within a neural network that modifies data using weighting factors for analysis purposes. The output of the hidden nodes creates the values for the output nodes (Dawson & Wilby, 1998).

Independent Component Analysis: A computational methodology similar to principal component analysis used to minimize the variance around an orthogonal vector with the data set (Hyvärinen & Oja, 2000).

Infit Statistics: Inlier sensitive fit statistic (Petridou & Williams, 2010).

Input Nodes: The node within the artificial neural network responsible for transmitting data to the hidden nodes (Somoza, Eugene & Somoza, 1993).

IRT True-Score Method: A method for determining population parameters from the item response parameters, discrimination (*a),* difficulty (*b*), and guessing (*c*), using Item Response Theory (Dimitrov, 2010).

Item Response Theory (IRT): Also known as latent trait theory, this theory is a modern test theory, in which, the true-score is defined by a subject's ability associated with a given trait on a logit scale (Hambleton & Jones, 1993).

Kaiser-Meyer-Olkin Measure of Sampling: A statistical method for examining the appropriateness of factor analysis by testing the underlying assumptions related to the sample (Stewart, 1981).

Latent Trait Reliability Method: A method used to determine inter-item reliability not dependent upon the assumptions of Cronbach's alpha (Dimitrov, 2009).

Least Square Distance Model: Is the minimization of matrix norms using the Euclidean least square distance (Glunt, Hayden, Hong & Wells, 1990).

Level 4 Biosafety Laboratory: The maximum levels of biological containment for infectious agents (Hawley & Eitzen, 2001).

Linear Learning: The assumption that learning occurs incrementally, in ordered discreet units from beginning to end (Grefenstette, Ramsey & Schiltz, 1990).

Linear Logistic Test Model: A method used to examine the validity if item constructs (Embretson & Gorin, 2001).

Logical Reasoning: A cognitive attribute used to develop a rational, systematic series of steps to arrive at a conclusion (Baril, Cunningham, Fordham, Gardner & Wolcott, 1998).

Mental Calculation: A cognitive attribute used to arrive at arithmetic calculation without the aid of computers, calculators or other external devices for numerical manipulation (Heid & Blume, 2008).

MERCI Model: A model of assessment replacing personal ability with attribute probabilities to measure a subject's cognitive attribute profile using real-time adaptive testing (Lamb, 2011).

Orthogonal Factors: Assumes that the factor loadings are uncorrelated (DeYoung, 2006).

Outfit Statistics: The outlier-sensitivity fit statistic (Petridou & Williams, 2010).

Output Node: The reflection of the input node once the data has been processed through the hidden node and weighted (Somoza, Eugene & Somoza, 1993).

P-16 Student: The grade levels ranging from pre-school (P) to the end of the 4-year bachelor's degree (16).

Parameter a: Item Discrimination (Petridou & Williams, 2010).

Parameter b: Subject Ability (Petridou & Williams, 2010).

Parity Judgment: A cognitive attribute used to estimate the equality between two quantities (Allik, Tuulmets, & Vos, 1991).

Perceptive Learning Rule: The learning algorithm contained within the artificial neural network (Carpenter, 1989).

Perceptual Binding: A cognitive attribute used to couple characteristics between items (Mitchell, Johnson, Raye, Mather, & Esposito, 2000).

Person Item Map: A graphical distribution of person ability versus item difficulty on a logit scale (Tesio, 2003).

Posteriori: Relating to or denoting conclusions derived by reasoning from of observed facts (Gauvain & Chin-Hui, 1994).

Principal Component Analysis of Residuals: A data transformation that uses orthogonal vectors through a data matrix, this approach accounts for the maximum amount of variance possible in the set of data (Rao, 1964).

Principle Factor Method: A method of factor extraction and is the first phase of exploratory factor analysis (Pruzek, 2005).

Projection Pursuit: A statistical method in which, one identifies deviations from a normal distribution in 3-D space (Li, 1991).

Q-matrix: A mathematical representation linking cognitive attributes with specific tasks using a one and zero to designate the presence or absence of an attribute (Choi, 2010).

Radial Basis Function: An activation function for a linear combination within a neural network (Specht, 1991).

Science Process: Processes that use deductive reasoning to produce empirically consistent results obtained through experimentation (Hodson, 1985).

Serious Educational Game: Serious Educational Games are a generalized form of Serious Games to mean games for purposes other than entertainment but that also specifically include training and educational purpose and pedagogy within their design (Annetta, 2008).

Sublexical Routing: A process in the subject converts a portion of a word, into its holistic form when reading, for cognitive processing purposes and ultimately comprehension (Au-Young, James & Howell, 2002).

Task-Attribute Relationship: The assignment of specific cognitive attributes required to complete a task. Psychologists outline this relationship within a Q-matrix (Gierl, Alves, & Majeau, 2010).

Theta (Θ): The mathematical representation of subject ability within the IRT framework used for comparative and analysis purposes (Petridou & Williams, 2010).

Verbal Production: A cognitive attribute responsible for the creation and processing of verbal language for comprehension purposes (Bock & Levelt, 2002).

Visual-Spatial Thinking: A cognitive attribute used to determine the position of an object in three-dimensional spaces (Kozhevnikov, Kosslyn, & Shepard, 2006).

Vygotskian Framework: An overarching theory of learning in which the individual constructs knowledge based upon the interplay of internal representations and external social interaction (John-Steiner & Mahn, 1996).

Zone of Proximal Development (ZPD): The difference between the tasks a learner can complete independently and those tasks the learner can complete with minimal assistance from a peer at approximately the same level of ability (Rezaee & Azizi, 2012).

CHAPTER TWO

Literature Review

This section provides an overview of the current literature relating to Serious Educational Games and their use in the classroom. The relationship of Serious Educational Games to education ties to increases in computational power and task authenticity. These increases lend to the potential development of cognitive diagnostic approaches using Serious Educational Games as an assessment platform.

History of Serious Educational Games and a Definition

Although there are overlapping domains related to SEGs such as e-learning, edutainment, and digital game-based learning, this study focused on the domains related to science education. Each component has specific characteristics and conceptions of how the learner interacts within the particular virtual environment. E-learning environments are a broadly inclusive category to include computer based training, online education and computer aided instruction (Bernard, Abrami, Borokhovski, Wade, Tamim, Surkes, & Bethel, 2009). This form of learning is typically isolated to the individual used with little to no synchronous interaction with the instructor or other students. Using Charsky (2010) as a basis, Edutainment is electronic forms of learning that are design to entertain and educate. Edutainment tends to be one-way and without interaction. An example of Edutainment would be an educational television show or documentary. Game-based

learning is a form of learning though game playing (Papastergiou, 2009). This form of learning can use video games and other electronic means to learn but is not limited to digital or electron play. It is through this interaction, between the environment and the student that results in learning. E-learning is a generalization of this computer-enhanced learning and distance-education approach (Demetriadis & Pombortsis, 2007). Following the increased use of E-learning in the 1990s, and its combination with new multi-media technology, Edutainment, or educational entertainment developed (Michael & Chen, 2005). However, more recently, edutainment has become associated with video games with learning intent. Researchers design edutainment approaches to target preschooler and younger children (Sarama & Clements, 2002). Creators of the Edutainment genre target this younger age group in an effort to expose them to science and mathematics, thinking and processing. Over time, the development of Edutainment stalled due to content associated with the games being thought of as "boring, drill, and kill learning."

Overall, one serious problem within the gaming industry was the lack of hardware development (processing power, graphic rendering, and interface development) in allowing for the realistic settings, tool interactions, and tasks to effectively create diagnostic educational games. Many of these limitations changed during the early millennia when individual processing power reached a sufficient level to make realistic 3-dimensional (3-D) renderings of environments possible (Yang, Tong, Yip, & Xu, 2009). This increase in processing power also coincided with new memory formats allowing the average user to have access to unprecedented quantities of computer memory. The outcome of this increase in processing and memory capacity was the evolution of the v-

learning environment or virtual learning environment from the e-leaning environment; these learning environments particularly focus on the K-16 teaching and learning (Annetta, Folta, & Klesath, 2010).

In response to this increasing memory and more realistic 3-D rendering capability, the United States Army released a game titled *America's Army* in 2002 (Gudmundse, 2006). To describe these new genera of games, Zyda and Falstein (2002), independently coined the term Serious Game (Newsome & Lewis, 2011). The release of the Army's Serious Game, in conjunction with the Woodrow Wilson Center introduction of the *Serious Games Initiative*, created the impetus within the educational sector to develop games for more than just entertainment, and decidedly placed the term *Serious Game* into the public lexicon. Through this initiative by the Woodrow Wilson Center to design and produce Serious Games, coupled with work by researcher-educators such as Annetta (2008), to add pedagogical and learning aspects to Serious Games, the term Serious Educational Game came about. Annetta conceptualized the term SEGs as a generalized form of Serious Games to mean games for purposes other than entertainment but that specifically include training and educational purposes within their design.

Out of the Serious Games drive, comes the concept of game-based learning. This particular branch of educational gaming or game based learning, deals with a very specific approach in which one defines learning outcomes (Kim, Park, & Baek, 2009). Game-based learning, the direct predecessor to SEGs, specifically has the capacity to enhance training, learning, and practice (Hayes & Games, 2008). From this fringe, the term Serious Educational Games has matured and current searches using Google Scholar

(03/2013) with the search term *Serious Educational Games* yields over 298,000 peer-review journal articles across numerous disciplines ranging from, education, computer science, business, and economics to engineering, natural sciences, and communications. Through its maturation within the literature, the term (Serious Educational Games) more recently has become more specific; referring to games designed to run on personal computers or video game consoles with the intent of training, simulating, and educating the subject, specifically targeting P- 20 content areas (Annetta, 2010). These games are designed specifically to take advantage of the engaging nature of video games through the bridging of cognition and psychological reward systems through stimulation (activation) of the areas of the brain associated with attention and arousal – namely- the frontostriatal and frontolimbic regions of the brain (Schmitz, Rubia, van Amelsvoort, Daly, Smith, & Murphy, 2008). A comparison of the SEGs with their non-Serious counterparts may provide the most telling definition of what a SEG is. Annetta and others suggest that while both SEGs and non-Serious Games contain art, story, development and software, it is the addition of the content and pedagogy, which separate the two (Annetta, 2008; Breuer & Bente, 2010; Maher, 2011). More specifically, the pedagogy of the task completion processes, and the learning content integration play the critical role while, story, and characters, etc. support these components and act as a means to promote affect arousal within subjects taking part in the SEG. (Johnson, Rickel, & Lester, 2000).

Games in the educational setting. The use of video games in the educational setting has been in existence for a significant period. Numerous empirical studies suggest that there is significant educational value to their use (Annetta, 2008; Annetta, Minogue, Holmes, Cheng, 2009; Mitchell & Savill-Smith, 2004). SEGs present the learner with complex representations of real-world problems within the educational environments. These complex representations would not otherwise be possible for a student to interact within the real world (Dondinger, 2007). For example, is very unlikely that a P-16 student would have access to or engage in learning within a Level 4 biosafety laboratory. The learner within these environments, video games, is exposed to complex representations often requiring specific tasks to be completed in order to forward the game toward the objective and, by extension, promote learning. Through task completion within the game, knowledge construction takes place and the video game acts as the mediator.

The construction of learning in a virtual environment is analogous to construction within other environments. This occurs because humans construct and use knowledge to identify and understand critical processes regardless of the environment. Thus, this construction is common while designing SEGs (Jamaludin, Chee, & Mei Lin Ho, 2009). The student develops concepts associated with learning through the generation and use of internal representations of concrete objects in the real world while using the virtual equivalent (Perlovsky, 2007). Thus, there is a tendency to focus on the faculties that develop recognition of the significant objects within a problem and solve for those objects (i.e. inferential and critical reasoning). Based in this understanding, one can

propose that computer game designers would need highly organized cognitive structures to facilitate internal representation. Therefore, it is plausible that these internal representations would be necessary in order to use science knowledge, when confronted with 'game situations'. Studies suggest that video game designers, and by extension SEG designers, would have the need to encode explicit information presented in the game for use later in task based problem-solving, thereby potentially transferring awareness and knowledge application to similar environments within the real world (Moreno-Ger, Burgos, Martinez-Ortiz, Sierra, & Fernandez-Manjon, 2007). This explicit encoding or knowledge construction, and knowledge deployment, is the key feature for the measurement of cognitive attribute sets. In other words, task completion is a key consideration when assessing cognitive attributes (Hadwin, Winne, & Nesbit, 2005). However, skill transfer across multiple domains and generalization of these cognitive attributes outside of the particular context of SEGs is still an area of intensive research (Baden, 2008). Specifically, the identification of patterns of cognitive processes used by SEG designers in multiple science domains is of critical significance to the science education and psychology community.

The primary assumption when exposing a student to science-based, educational computer environments is that the students (subjects) undertake specific tasks when using the SEGs and the tasks result in learning gains. This assumption is the underlying principal of educational gaming, more specifically SEGs (Annetta et. al, 2008). However, in many cases, educational software incorporates design principles that mimic recreational software as a means to engage students (Thomas & Macredie, 1994). One

particular domain of educational software that illustrates this point is science based SEGs. In many cases, SEGs purposefully imitate their recreational counterparts with the hope of creating engagement in tasks leading to learning. This engagement is also a key point in the extension of SEGs as a means to measure student performance.

Alternative Means of Measurement and Assessment in Education

Item Response Theory (IRT) and Classical Test Theory (CTT) provide a means for obtaining examinee's scores and scale measures for latent traits (Harvey, 1999; Lamb, Annetta, Meldrum, & Vallett, 2011). However, it is not possible to use IRT and CTT to assess and profile combinations of latent traits and attributes (Henson, Roussos, Douglas, & He, 2008). An alternative approach to these measurement techniques (IRT and CTT) is the estimation of the cognitive attributes mastery patterns (CAMPs) and Cognitive Diagnostic Assessment (CDA) models. Researchers can develop each of these measurement techniques (CAMPs and CDA) through cognitive diagnostic approaches originally developed in clinical psychology. While educational testing data can provide meaningful information to guide student learning, there is still a need to develop an understanding of the underlying psychological processes, which can help to model and explain underlying testing outcomes.

CAMPs and CDA, diagnostic models are appropriate when the educational focus is not the estimation of general student ability (Θ), but the estimation of specific task ability (A) (Almond, DiBello, Moulder, & Zapata-Rivera, 2007). More specifically, these models provide for the probability of correctly answering an item or completing a task as a function of a particular cognitive matrix pattern or attribute mastery pattern. These

models can provide a means to access latent attribute profiles of student cognitive strengths and weaknesses (Gierl, Cui, & Zhou, 2009). Researchers within the field of cognitive diagnostics seek to develop, the power of the CDA approach, not at the district level as it is currently employed but at the individual student (classroom) level (Roberts & Gierl, 2010).

Despite the potential power of student-level, cognitive diagnostic assessment (CDA) reports, there are limitations to current reporting methods for CDAs. Typical CDA reporting methods center on the aggregation of large numbers of samples at the national, state, or district level (Hattie & Jaeger, 1998). At these levels of reporting there are often questions of timeliness, and meaningfulness to the classroom level practitioner. This disconnect between levels of reporting can potentially be mitigated with the use of computerized testing and direct access to task-based assessments via virtual tasks. Computerized testing provides a means for educators to develop large data streams for analysis beyond content outcomes. Task-based assessments using cognitive diagnostics can also aid educators in their understanding of student learning. In particular, a thorough review of these assessments can allow for a view of student cognition while engaged in the learning process.

Cognition

Psychologists see individual human cognition as a means to embody a set of processes and mechanisms by which an individual understands the world though thinking and problem solving (Wilson & Keilm, 2001). Brown defines general cognition as, a combination of several cognitive attributes activated in parallel and simultaneously. This

process of understanding has in some ways led to an artificial separation between the knowing and doing (Brown, Collins, & Duguid, 1988; Hotton & Yoshimi, 2011). One-way to bridge this separation between knowing and doing is with authentic tasks. Authenticity is a key feature of learning and by extension learning in the SEG environment. The authenticity found in SEGs derives from the realistic immersive environments rendered via the design process. Thus, there is a linkage between the product of learning and the structure of the activity. This linkage between the structure of the activity and the authenticity demands that the one's processes (cognition) situate into the context of the learning environment. Within the context of the situated cognition, most adults have a system of skills that are rooted in their perceptual process that develops throughout their early life and into their late teens. These processes develop into a set of diverse and complex cognitive procedures when used in parallel and simultaneously (Langley, Laird, & Rogers, 2009; Moreau, 2012). These component skills or attributes include, but are not limited to, the ability to comprehend and produce written and oral statements describing the interaction of complex variables, critical reasoning, and the ability to retrieve, calculate, estimate and reason through simple and complex problems. These relationships result in numerical and mathematical conceptual and procedural integration of activities (Wang, Geng, Hu, Du, & Chen, 2013). Given the connection between mathematics and science, portions of these task activations of cognitive attributes also activate when engaging in science processing (Hestenes, 2010).

Conceptualization of the underlying cognitive architecture in mathematical and science processing skills is the focus of intensive research and discussion (Kalyuga,

Rikers, Slava, & Paas, 2012; Kroeger, Brown, & O'Brian, 2012). First, the discussion centers on the extent to which the distinct attributes exhibit local independence. Second, there is disagreement around the specific degree to which mathematics and science processing are specific to individual encodings. Several researchers note the formulation of encodings found in the processing of science skills related to quantified magnitudes (Kolkman, Hoijtink, Kroesbergen, & Leseman, 2013), verbal production (Deans, 2010), inference (Alfieri, Brooks, Albrich & Tenenbaum, 2011), reading (Ozuru, Dempsey & McNamara, 2009), and parity judgment (Gobel & Snowling, 2010). Such phenomena indicate that functionally distinct processes activate in parallel and interact, because of different functions involved shared representations. Further review of attribute activation profiles suggests a hierarchical relationship between the attributes. Specifically, vision-spatial representations are integral to magnitude judgments, estimations, and arithmetic (Green, Feinerer, & Burman, 2013). Similarly, verbal-code structures encode fact retrieval processes. Within current models, there are assumptions that comprehension processes and interaction between attributes convert different surface forms into common abstract formats for input and storage. The cognitive architecture discussed in this study is broken into seven domains for exploratory and explanatory purposes. These domains act as a means to structure and organize the cognitive attributes. The seven domains are attention, mathematical learning and deductive reasoning, inductive reasoning and decision-making, perception, language processing and action. To organize the cognitive attributes further, the study assigned the attributes to domains based upon activation during task completion. The domains are Attention, Mathematical Learning and

Deductive Reasoning, Inductive Reasoning and Decision Making, Perception, Language Processing and Action.

Cognition related to the domain of attention. Cognition provides a framework for clarity and, each of the attributes rests upon a framework bridging between neuroscience, science education, and educational psychology. Specifically, the following portion of this study develops both the conceptual and operational definitions, which drive the data mining and exploratory nature of the study. The definitions of attention switching and visual attention are rooted in the more generalized process of selective attention. Selective attention generally refers to a set of operations, which assist in the determination of how to analyze multiple input streams (Song, 2011). Due to the limited processing capability associated with cognitively intensive tasks, it becomes necessary to allocate cognitive resources in a selective manner. The first of these possible modes for discrimination is attention switching (Hanania & Smith, 2010) and the second is visual attention. Mayer, Roebroeck, Mauer and Linden (2010), define attention switching as the subject's ability to change between performing multiple, individual tasks. Subjects experience activation in the areas associated with the dorsal posterior parietal and frontal cortex with transient activation in the occipital region (Corbetta & Shulman, 2002; Pennick & Kana, 2012). Psychologists define visual attention as the development of a scene through a combination of attention, eye movement, and memory. Subjects activating the visual attention cognitive attribute experience signaling within the visual cortex and superior colliculus (Posner, Peterson, Fox, & Raichle, 1988; Posner, 1990; Pasqualotto & Proulx, 2012). Peripheral stimuli localization is an outward expression of a

subject's attention, and is often related to visual attention (Hoyer, Cerella, & Buchler, 2011). Visual attention is thought to act as the initial stimulus drawing the subject's attention followed by peripheral localization (Hubert-Wallander, Green, & Bavelier, 2011).

Cognition related to the domain of mathematical learning and deductive reasoning. Mental calculation (arithmetic numerical sense), is another component suspected of being used within the science process and is defined as the ability to determine the exact size or quantity of a system. Activation in the areas associated with this attribute is located in the left temporal lobe (Moeller, Wood, Doppelmayr, & Nuerk, 2010). Another related cognitive attribute is mental calculation; the subject uses mental calculation or the ability to complete mathematical operations without the aid of external devices in estimation of numeral attributes. These estimations when conducted in parallel with one another determine the ability to approximate size or quantity. Functional MRI studies note that activation of the parietal cortex occurs during estimation tasks (Dormal, Dormal, Joassin, & Pesenti, (2012). A third related attribute is that of quantification; this attribute is the mental processes associated with counting, and measuring as function of inputs from the senses (Berglund, Rossi, & Wallard, 2012). The areas of the brain activated when engaged in quantification, are the dorsolateral prefrontal lobe, and the temporal lobe (Arsalidou & Taylor, 2011). Acting as a subcomponent of quantification and mental calculation is the attribute parity judgment, or the ability to judge whether two quantities, verbally or in written modes are equal. Researchers observe activation of the

intraparietal silcus and the superior colliculus when subjects complete tasks related to parity judgment (Winkler, 2009).

Cognition related to the domain of inductive reasoning and decision-making. Critical reasoning or logical reasoning is the formation of conclusions, inferences, or judgments based upon stimulus input. Science educators suggest that critical reasoning is a key or even the key attribute used in the science learning process (Bond, Philo, & Shipton, 2011). Educators explain critical reasoning as a mental process and procedural moves culminating in a logically valid conclusion (Manktedlow, 2012). Functional Magnetic Resonance Imaging (fMRI) studies of reasoning tasks result in activation centering in the striatum area of the brain (Barbey, Koenigs, & Grafman, 2012). Inference is a closely related attribute to critical or logical reasoning. Inference is a process of deriving logical conclusions from known premises (Hayes, Heit, & Swendsen, 2010). The more generalized attribute of critical reasoning incorporates three components of perception, expectation, and inference (Ren, Schweizer, & Xu, 2013). This relationship between critical reasoning and inference provides a view of the hierarchical nature of cognitive attributes and helps to explain the differential weightings found in artificial neural network models. The final cognitive attribute found within this domain is that of memory retrieval. Memory retrieval is the process by which one accesses a stored memory and activates it as part of several cognitive interactions such as recall, recognition, and remembering (Cabeza & Moscovitch, 2013). Areas of the brain associated with the retrieval are located in the temporal lobe, the amygdala, the

hippocampus, the rhinal cortex and the prefrontal cortex (Ku, Tolias, Logothetis, & Goense, 2011).

Cognition related to the domain of perception. The fourth domain associated with the cognitive attributes identified within the interaction of science process is that of perception and perceptual binding (Teufel, Fletcher, & Davis, 2010). A definition of perceptual binding is the ability to link characteristics between multiple items an individual perceives within the environment (Mance, & Vogel, 2013). While perception is similar to pattern recognition, the key difference is pattern recognition results from an a priori approach, while perceptual binding is posteriori (Linhares, Freitas, Mendes, & Silva, 2012). Neuroscientists see activation of the neocortex when studying subjects engaging in perceptual binding tasks. A second cognitive attribute located under the domain of perception is that of spatial ability. Spatial ability is the attribute associated with the perceptions of the visual world in a 3-D environment. In addition to the perceptions associated with the world, it is also important for the recreation, transformation, and modification of visual aspects of experiences (Hoffman & Nadelson, 2010). Kravitz, Saleem, Baker and Mishkin,(2011) observed activations during tasks related to spatial ability in the hippocampus, posterior parietal cortex, entorhinal cortex, prefrontal cortex, retrosplenial cortex, and the perirhinal cortex .

Cognition related to the domain of language processing. Reading is the means by which individuals link orthographic symbols to phonological, sematic, morphological, and grammatical information to develop meaning (Rastle, 2012). Reading experts suggest the cognitive attribute reading, relates to other attributes such as sublexical routing and word recognition. Areas of the brain associated with the reading attribute are associated with the frontal lobe, the occipital lobe, and temporal lobe (Buchweitz, Mason, Tomitch, & Just, 2009). A second cognitive attribute used within the science process is that of verbal fluency. Verbal fluency is the ability to rapidly, mentally-access vocabulary while speaking and writing (Birn, Kenworthy, Case, Caravella, Jones, Bandettini, & Martin, 2010). Subjects activate portions of the brain during the use the verbal fluency. These portions of the brain are the frontal lobe, temporal lobe, parietal lobe, occipital lobe and the limbic lobe (Binney, Embleton, Jefferies, Parker, & Ralph, 2010).

Cognition related to the domain of action. The smallest unit of control within the domain of action is the motor unit. The motor unit consists of a synaptic junction, a motor axon, and the associated muscle fibers. Recruitment of the motor units occurs under the motor cortex of the brain (prefrontal cortex) through integration of sensory inputs (Neary, 1997). Motor control is itself the function of supervising motor activities from a cognitive perspective. Researchers have isolated motor activities to the motor cortex portion of the brain (Reynolds, Lane, & Richards, 2010). Related to the motor control attribute are, motor after effect, motor execution, motor inhibition, motor planning, motor program, and sequencing.

Cognition and video games. Teaching methods in the physical sciences, specifically chemistry, have traditionally employed models in a two-dimensional format as illustrations within textbooks. Illustrations assist in the spatial-learning and memory of chemical structures. Several studies suggest virtual manipulatives, simulations, and SEGs positively affect student achievement (Lamb & Annetta, 2009; Criswell, 2011; Tolentino, Birchfield, Megowan-Romanowicz, Johnson-Glenberg, Kelliher, & Martinez, 2009). Furlan and Bell-Loncella (2010) also suggest that the combination of computation and visualization in the form of modeling software, via SEGs, improves student understanding of chemistry concepts. The benefits of chemistry models help the learner's understanding of the arrangements of molecules in solid matter in a variety of settings. In the examination of the efficacy of various types of modeling and visualization for chemistry learning, results have indicated that three-dimensional representations, in particular those representations in a SEG environment, better supported student understanding of molecular structure and resulted in greater student enthusiasm for learning the tasks (Lamb & Annetta, 2009; Limniou, Roberts, & Papadopoulos, 2008). Other studies demonstrate that learners are more apt to perform better using visual displays that educators have optimized to reduce cognitive load. Wang & Borrow (2011), building on Wu and Shah (2004), have suggested that visual-spatial thinking is an important aspect to successful learning of chemistry concepts. Literature linking cognition to the specific types of learning is a natural outcome of video game design for learning and education. However, a consistent finding regarding the relationship of cognition to video game design is the transference and improvement of multiple cognitive

attributes. Specifically, transference of peripheral localization, visual attention, attention switching, general cognition, visual-motor coordination, and spatial ability develop rapidly from video game design and use (Li, 2009; Spence & Feng, 2010). Each of these areas in particular have been shown to have a relationship to Science Technology Engineering and Mathematics (STEM) fields such as architecture, engineering, and drafting along with more generalized fields such as piloting, mechanics and machine operation (Uttal & Cohen, 2012).

From a neurochemical point of view, successful video game design also stimulates cognition though the release of dopamine. The stimulation of dopamine, as a model for learning, provides a linkage from biological response to psychological outcomes making learning via video game play one of the best mapped out phenomena in education (Waldmann, 2012). By providing an underlying biological reasoning for educational outcomes, educators can more specifically target interventions understood to derive from biological (neurochemical) means. Dopamine acts as one of the many neurotransmitters allowing for the modulation of information transference between one area of the brain and another. This is important when considering the transference of virtual tasks to task within the real-world. Multiple cognitive attributes aid transference through executive functions triggered via the release of neurotransmitters such as dopamine. Positron Emission Tomography (PET) illustrates the release of significant amounts of dopamine from the brain during successful video game design and the resulting activation within areas of the brain associated with the executive functions. This release is particular closely tied to the areas tied to reward and learning (Bateman &

Nacke, 2010), thus providing a means to explain persistence and fluency found while subjects design SEGs. This release seems to play a crucial role in learning by stimulating the reorganization of neural pathways.

Cognitive Diagnostics

Nationwide, our schools have increased the number of student achievement and content tests; this increase results from the push for data driven decision-making (D3M) approaches to student learning. This increased focus, on measuring students' abilities for placement into the appropriate grade and level for their educational needs, has led to concerns from educators such as overreliance on test outcomes and increased accountability for educators for outcomes not directly under their control. However, educators have always been challenged by attempts to measure indirectly observable characteristics, such as efficacy (Bandura, 2006), computational thinking (Qui, 2008), and other internal, latent constructs. Measurement of these latent attributes is contingent upon observed responses to items indicating attributes (Lord & Novick, 1968; Wang & Chang, 2008). This reasoning is analogous to the difficulty educators have in identifying underlying cognition when students engage in the learning process. A means to discover these underlying processes is through CDA models and by extension the Q-matrix. Researchers often link attributes and items using a Q-matrix (Von Daver, 2010). One area of exploration is the use of SEGs to present tasks related to particular attributes (Kirriemuir & Mcfarlane, 2004).

Researchers from the beginning of the testing movement have posited that subject performances on specific test items (tasks) are contingent upon specific cognitive aspects

called attributes. For the purposes of the study, the term cognitive attribute is a skill or process that a subject must possess to solve a particular task or item (Gierl, 2007). The Q-matrix is a mathematical model associated with the cognitive diagnostic methods. Educational measurement experts classify individual responses into categories of cognitive attribute patterns based upon test or task performance (Sventina, Gorin & Tatsuoka, 2011). Classification of response patterns into the Q-matrix depends upon the estimation of a subject's ability (Θ) and item difficulty (b) (Briggs & Alonzo, 2012). Using the specific items (tasks) linked to the necessary cognitive attributes (in a conjunctive model) that make up the assessment; one can develop patterns seen as acting as an ideal response pattern (task completion pattern) for a specific knowledge state. The results from the response pattern analysis develop into likelihoods of correctly completing the tasks. Through the linkage of testing and cognitive attributes, the fields of cognitive psychology and psychometrics are able to bridge (Dimitrov, 2007; Kaufman, 2011). Dimitrov uses the least square distance model (LSDM) to evaluate the probability of an item being answered corrected using an IRT model (Dimitrov, 2007). The equation below specifies the calculation of the attribute. The two models commonly used for the development of the cognitive diagnosis are the conjunctive model and the disjunctive model. Within the conjunctive model, each attribute is a key component and necessary for task completion. Specifically within the attribute mastery pattern (AMP), removal of one attribute from the pattern results in the subject failing to complete the task. In other words, each attribute is critical to task completion. This model assumes the requirement of local independence for the attributes. A second advantage of this model is the ability to

frame the AMP as an IRT measurement problem. The ability to analyze the attribute mastery patterns as IRT problems provides a significant advantage in that cognitive attributes are on a common scale of logits for comparison purposes. Equation 1 provides the means for the calculation of the probability of success in completing a task under an IRT model.

$$\text{Pi}= \prod_{i=1}^{k}[P(Ai = 1|\Theta)] \quad (1)$$

Learning Theory

Laboratory based learning in the sciences has been described as a series of isolated skills, developed in linear fashion, resulting in successful processing of tasks for completion in the form of learning progressions (Hipkins & Kenneally, 2003). More recently, studies of laboratory science suggest that this approach, isolating specific skills, results in a poor success rate for completion of the laboratory and content learning (Songer, Kelcey & Gotwals, 2009). Psychologists and educators attribute this lack of success using these methods to the non-linear aspects of learning, thus measures that rely on the linearity of the learning as a means of progressing may not be adequate to model students' learning. Thus, Bayesian approaches such as ANNs may be of significant use to researchers. In this framework, full assessments, which integrate learning, such as that, found in SEG environments, where the game presents all skills and actions in an open-ended, non-linear format reflects learning behaviors found in real-world settings.

Further evidence of non-linear learning is in the process of equilibration. Equilibration of new learning experiences via assimilation and accommodation takes

place in a non-linear order. The non-linear presentation of tasks is present in SEGs via the open-ended formatting and interactions. This lack of linearity in learning can confound currently available assessments as current curriculums assume a linear progression of topics (Xu, Meyer & Morgan, 2006). The misapplication of information and process occurs due to the lack of exposure to the overall big-picture view of the place of the task in the science process. Within the traditional linear format, contradictions and confusion continues until the learner's mental concepts stabilize via successful application of disparate information, further developed into meaningful knowledge and application via appropriate cognitive attributes (Clements & Samama, 2011). Equilibration or reconstruction of conceptions is often a self-regulated process mediated through affective and cognitive processes (Volet, Vauras, & Salonen, 2009; Kitsantas & Zimmerman, 2009).

SEGs by design exemplify scaffold type learning using Zone of Proximal Development (ZPD). The ZPD is a point, which resides on the edge of a person's understanding related to a concept. Incorporation of this knowledge is often contingent upon meaning making associate with abstract information represented symbolically. It is due to the symbolical nature of the conceptual representation that links the symbols of the SEG with the symbols of the real world. Based in this view, one can suggest that learning within the SEG takes place under a Vygotskian framework. Vygotsky views learning in terms of symbolic representations via language and abstractions (Vygotsky, 1978). New material presented below the ZPD results in low-level learning and little stimulation of cognition. New material presented above the ZPD results in confusion and cognitive

overloading evidenced by mistakes and lack of task completion (Fayol, Largy, & Lemaire, 1994). Students engaging in learning compare new and old conceptions and evaluate usable heuristics to solve novel problems at the critical point within their ZPD. Item-task misfit analysis reveals low-level and confused learning outcomes when evaluating items (tasks) using the IRT framework. Thus, with the use of IRT, it is possible to quantify ZPD misfit.

The linkage of cognition (higher-order) and more traditional models of learning in the form of computational linkage illustrated through ANN models arises out of attempts to develop artificial intelligence (Efendigil, Onut, & Kahrman, 2009). Artificial Intelligence treats higher-order thinking as a computational task. Due to the treatment of higher order thinking as a computational task, the cognitive attributes identification allows for their parameterization via IRT. Researchers contrast this view (modern computational view) of cognition with more traditional views of cognition in which higher-order thinking is treated as a symbolic and representative endeavor (Woelert, 2012). Bruner also reflects these assumptions that human cognition is symbolic and operates within an externalized, sequenced linear progression (Bruner, Goodnow, & Austin, 1986). This view culminates in the conceptualization that the human mind and brain along with computers act as a physical symbolic system operated upon by serial processors using memory and language (Lipinski, Sandamirskaya, & Schoner, 2009). A concern when viewing learning as a symbolic model is the limitations associated with current models, using traditional assessment data sets within the representation of internal processing with novel information, without known algorithms and rules. These

limitations make it difficult to model the human process of acquiring information and problem solving in the context of science processing.

Pilot Study

This section of the dissertation discusses a pilot study conducted during the spring of 2012. The purpose of this pilot study was to provide evidence for the validity of techniques used to integrate conceptually Bayesian Models (Artificial Neural Networks) with Item Response Models and cognitive diagnostics. This section also provides the background information needed to understand the descriptions and results of the piloted techniques.

Pilot study description. The purpose of the pilot study was to establish an exploratory procedure for understanding the role of latent cognitive attributes associated with the use of SEGs play to teach science processing within the context of biotechnology. A secondary purpose of the study was to uncover the associated cognitive attributes used while using a science based SEG. Lastly, the study seeks to develop a means to create an artificial neural network (ANN) model of the interaction between the cognitive attributes and game task items. The research questions addressed within the pilot study were:

RQ1. What are the underlying factors exhibited through the measurement of task items associated with Serious Educational Games?

RQ2. What are the cognitive attributes that underlie Serious Educational Game play?

Consideration of the research questions and the supporting literature suggests the following hypotheses: while playing SEGs within immersive learning environments, it is possible to map the relationships between items, factors, and cognitive attributes using a cognitive diagnostics approach through development of an artificial neural network model.

Pilot Study Method

The purpose of this section is to outline the methods initially used to develop techniques refined in this study. This section primarily focuses on the use of Factor Analysis and Artificial Neural Networks to establish cognitive relationships.

Pilot study sample. This study analyzed data from 500 students located within multiple states within the Southeastern and Midwestern portions of the United States. Subjects' grade levels ranged from 9th grade to 12th grade; using grade as a proxy for age, the subjects' age range from 14 to 18. Data collection occurred via real-time, server, data logging, as the subjects played Mission Biotech (MBt). MBt is a first person SEG designed to teach concepts related to biotechnology. Logged server data consisted of all actions taken while playing MBt. Examples of actions included; player non-player character interactions, tool use, and key strokes. The total number of analyzed actions was n=132,453 actions. Each study subject has a unique player identifier to assist in data sorting from the server logs. Each subject had prior exposure to some biotechnology education within their science course work and therefore possessed cursory domain knowledge but lacked task specific knowledge related to task completion within the

game. It is important to understand the intent of the study was not to measure subjects' gains in content knowledge, but to understand the cognitive attributes associated with the science based tasks the subjects completed.

Pilot study design. The study design was a one-group, posttest only design. Lack of pretest creates positive and negative outcomes for internal validity of the measure (game actions and tasks). Positive data manipulation outcomes (model fit) suggest that there are no threats to internal validity of the measure from pretesting. This lack of pretesting reduces the overall threat to internal validity of the task items by reducing item task familiarity and carry-over. However, without a proper control group it is difficult to assess changes due to treatment effects. Methodologists suggest this design for initial exploratory studies such as this pilot study help to establish initial study questions and hypotheses. Embedding the study intervention in the classroom creates a more true-to-life classroom experience and the quality of collected data may be superior.

Pilot study task presentation. Mission Biotech (MBt) is a first person SEG grounded in a problem-based learning model. MBt provides subjects with a realistic approximation of scientific problem solving and research work in a scenario, task-based presentation. Upon initial login, subjects receive a description of an outbreak of an unknown disease occurring somewhere within the world. As the introduction progresses the game assigns the subject's character to an elite organization known as Mission Biotech, which is responsible for attempting to stop the viral outbreak. The subject (player) is then required to complete various science-based, problem-solving tasks in a SEG environment to forward game play. The complex tasks involve, but are not limited

to, the use of polymerase chain reaction, DNA base-pair identification and other laboratory based actions and equipment use. Through these processes, the game leads the subject to identify and isolate the source of the viral outbreak while learning biological concepts through mentor characters, readings, and task-based training.

The subjects answer items and complete tasks in a drag and drop interface (matching), and a behavior-task outcome interface. The behavior-task interface is an interface in which the subject completes tasks designed to mimic closely their real-world counterpart. An example of this behavior task is balancing a centrifuge using blank centrifuge test tubes when spinning their samples. Completing the particular task items result in their progression through the game. The game logs each action (mouse clicks, inventory interactions, and questions answered, etc.) taken in the game, per session, per subject. Each subject action is then coded using a dichotomous outcome approach; coding was either a "1" for successful completion of the task item or a "0" for unsuccessful completion of the item or task. Through cumulative completion of the simple tasks, subjects are able to show mastery of complex tasks, subtasks, and content knowledge related to the game. The relationship allows for the suggestion of hierarchical associations between attributes. To address researcher bias data mining approaches were used to analyze emergent patterns within the data set and allow those patterns to guide the analysis and task-attribute relationships. In addition, the use of outside reviewers provides a guard against bias as the researcher did not assign task-attribute validity coding.

Pilot study results

A reasonable approach to determining the number of factors to fit the cognitive attribute is through a scree test (Cattell, 1966). The scree test plots the Eigenvalues in descending order and shows the point at which they level off. Results of the scree plot suggest that there are three underlying factors using the root ≥ 1 criterion (Dimitrov, 2010). Using root >1 criterion, eigenvalues less than one are not considered important because the variance that each standardized variable contributed to the extraction equals one. Methodologists suggest the root ≥ 1 criterion for 40 or fewer variables. Results of the scree test suggest that there are three underlying factors within the proposed construct. The outcome of the exploratory factor analysis support earlier results (eigenvalues) for the Varimax rotated simple solution. Analysis of the linear relationships and resulting correlation coefficients suggests that there may be a slight positive linear relationship between factor 1 and factor 2. Review of the results for factor 3 show a slightly negative relationship between factor 3 and the other 2 factors.

Results of the factor loading suggest there are three orthogonal factors present during game play. Post rotation of the three factors indicates, X1, X3, X5, X7, X8, and X10 relate to factor 1. Further examination shows that X2 and X6 relate to factor 2, X4, and X9 load on factor 3. Based upon this array and types of items on each factor, it is logical to suggest that factor 1 is game control, factor 2 is flow, and factor 3 is science processing knowledge. The three common factors game control interactions, flow, and science processing knowledge accounts for 67.48% of the variance in the observed variables.

Parameter estimates for a 1PLM model indicate adequate model fit. Fit statistics for the logit model suggest that items within the model are functioning within an adequate range. Item infit statistics range from 0.45 to 1.36 indicating proper functioning of the items and very little distortion associated with using the game tasks as a measure of each factor. However, analysis of the item outfit statistics suggests that items 4 and 5 are susceptible to outlier influences; the effects of these outliers are ameliorated by the large '*n*' (*n*=500). Item-measure difficulty results suggest that Item 4 provides the greatest level of difficulty and item 2 shows least difficulty.

Successful completion of X3 increases the probability of the successful task completion of the remaining items, as it is the most "difficult" item to complete. Conversely, X2 shows the greatest probability of successful completion. These results lend validity to the proposed model, as they align with the measure difficulty. Calculation of the mean parameter weights involved aggregation of the individual parameter weights resulting from the neural network analysis.

Fit statistics indicate that the research fails to disconfirm the IRT model as means for establishing probable task completion. Key to this point is the interpretation of theta (Θ). Within this proposed model, Θ represents the residual skill associated with the item attribute. Development of the hypothesized model allows for the use of the suggested artificial neural network model to develop propagation weightings. Probabilities for subject mastery of individual factors via APM show high variance accounting (r^2 = 0.9783). Parzen described the expected probability for each attribute via mathematical equation (Parzen, 1962).

Removing the Θ parameter from the Bayesian model (ANN) fosters the development of a discrete model, allowing for cross application of models, and the attributes of the developed model, to be more transferable for external model comparisons. The artificial neural network (ANN) model is most likely to propagate attribute 1 given appropriate stimuli for each cognitive attribute. Without the presence of Θ_i in the Bayesian Model, each item, *i*, the probability, π (Bayesian), and $\prod$ (IRT) have similar interpretations. The interpretation of each symbol represents the probability of solving an item given the subject has mastered the requisite skills.

Each weighting (coefficient) represents the strength of propagation through the neural network. A weighting of 1.71 for the factor 1 (F1) indicates that attribute 5 (A5) has the greatest impact on the probability of completion of items associated within the factor. A8 shows the least impact on item completion. Factor 2 (F2) task completions is most significantly impacted by A2 and least impacted by A7. However, one should note, that the differences between F2A4 and F2A7 propagation weightings are negligible. Factor 3 displays attributes from greatest impact to least-. Factor 3 consists of A8, A9 and A11. The comparison of F3A9 and F3A11 shows approximately the same propagation weighting.

Based upon factor analysis, item content, and expert review, the study assigned cognitive attributes to each factor. While each attribute listed reveals a positive weighting, there are negative weightings that act to counteract the attribute. However, it is beyond the scope of this pilot study and the proposed model to discuss in detail the effects of the negative weightings. For the purpose of the pilot study, negative weighting

are inhibitory to task completion when using those cognitive attributes. However, exploration of these attributes is beyond the scope of this pilot study.

The Q-matrix does not represent the definitive listing of factors and attributes for each of these tasks. Training the ANN to recognize the relationships between each attribute develops the mean probabilities (Modeled Bayesian Probabilities) for each factor within the Cognitive attribute matrix. Placing each group of attributes into the training model creates the recognition by the ANN of the cognitive attribute relationships. Upon completion of the ten thousand training iterations, the neural network calculated the probability of propagation using non-linear, neural network, propagation. Propagation probabilities and endorsement probability, in conjunction with parameter weightings, allow for the development of a modified Q-matrix. This necessitates the need to use dichotomous outcomes (1 = neuron does propagate, 0 = neuron does not propagate).

Pilot Study Discussion. The use of contrast-oriented design is important in the development of valid and usable CDA models. Tatsuoka (1983) established the construct-oriented design used to identify attributes that represent the skills and attributes, which when assessed provide meaningful information. The outcome of this process is the development of the Q-matrix tying the items tasks to the specific cognitive attributes. A cell within the matrix Q_{ij} takes the value of 1 if the mastery of the skill (k) is required to solve the represented items. The result of the study suggests that it is possible to use a science based SEG environment to generate an understanding of the underlying cognitive processes using a cognitive attribute matrix, forward-feed artificial neural network, and

classic factor analysis. Consideration of the results suggests rejection of the null hypothesis as adequate model fit and that generation of the Q-matrix is possible. The strength of the neural network analysis is that it allows for the analysis of the relationship between fewer items and attributes than in a traditional cognitive diagnostic analysis. Literature suggests that this result is congruent with expectations as convergence of the neural network model occurred after 1000 iterations.

The emergent factors developed from the factor analysis provide a starting point for the process of matrix development. The researchers classified game data into three categories: Control Interaction, Flow, and Science Process. The limited external instructions given to the subjects, and the repeated exploratory behaviors (subject repeating steps to ascertain cause and effect) seen within the data, illustrate that the subjects had no task-specific and little domain-specific knowledge. Factor analysis assesses the commonality of actions within the observed variables as part of a larger construct. For this study there appears to be three orthogonal factors (Research Question 1). Rotated solutions reveal a simple structure with three linearly independent factors (Thurstone, 1947). Cross-validation of the three factors corroborates the exploratory outcome. Each of these three factors results from extrinsically measured item loadings on the appropriate latent trait. Examples of items within the game control interactions construct are Help Menu Interactions and Resource Use and Interactions. Examples of the flow construct are Time on a Particular Task and Total Amount of Time. Rank and Badge Attainment exemplify the final construct, science process.

Parameterization of the three factors occurred using the 1PLM model (D=1.0). Infit, outfit and X^2 suggests adequate fit of item parameters. Model fit provides an adequate probability of action at the 95% confidence interval (Pilot Study Research Question 2). The researcher extracted post parameterization probabilities of subjects engaging in particular actions. From these extractions, artificial neural network weightings were developed. The study used artificial neural network weightings to identify the highest probability of action propagation throughout the network. The assignment of higher weighting, by the artificial neural network (ANN), results in those attributes belonging to the highest levels within an attribute hierarchy matrix and thus indicating upstream and downstream attributes. The neural network activates attributes with a higher possibility of propagation prior to activation of lower level attributes, thus creating a cognitive attribute hierarchy. This lends evidence to the appropriateness of the use of a forward-feed neural network and adequate neural network fit. This model of placement within a hierarchy also helps to explain why attributes -for example A5, which have a higher weighting, but also have a lower probability of completion. In this model, A5 is lower (downstream) than A1 in the hierarchy because A5, although it has a greater weight, it does not have a higher probability of propagation. Thus the order of consideration for ANN development should first be the probability of propagation and then secondly the weighting within the propagation. Neural network weighting, when placed in the factor loading positions, can also help to elucidate the relative 'pull' each attribute has within each factor assisting in the development of the hierarchy.

The study placed each attribute, based upon the weightings, and the extracted probabilities; taking into account the characteristics of the attribute description by the panel of experts. Based upon these characterizations, three experts in the field of cognition evaluated placement and naming of the attributes. A reliability coefficient of 0.70 between each reviewer suggests that weightings, and the context of the subject actions, support the attribute descriptions. One additional area not addressed in this study is the role of negative attributes. For the purposes of this study, negative attributes inhibit network propagation, or in other words, act as part of a backward propagation network. Lack of backward propagation does increase the probability of overfit error. This problem is in need of more study in order to develop this portion of the model. Tying the attribute descriptions to particular factors provides for the development of a cognitive profile of some cognitive attributes used within the SEG.

Evaluation of the factors, probability of endorsement, probability of propagation, and attribute descriptions, results in the development of a Q-matrix. Remembering that the Q-matrix relates the attributes to items through factors, consideration of the model results suggests that the model is an appropriate description of the CAMP (Pilot Study Research Question 3). Confirmation of the model can occur through appropriate structural equation models, which are beyond the score and breadth of this current study. Each factor acts as the extrinsic measure of the latent attributes. Through this reasoning, it is possible to represent Q-matrix graphically, wherein arrows from F1 to A5, F2 to A4 and F3 to A9 indicate that each of the attributes is subordinate to the larger factor domain. Examination of weightings shown in the Q-matrix showing that A1 and A10 are

subordinate to A5 as a set of cognitive operations required by the items associated with F1. This subordination supposes that the operation A1 and A10 require the operation of A5 a priori. Therefore, assessing the A5, A1 and A10 within matrix Q, show $Q11 = 1$ while $Q12 = 0$. Other entries within the Q-matrix show similar patterns.

Evaluation of the factors found in the Q-matrix indicates that subjects playing a SEG exhibit cognitive process similar to those found in a laboratory-based problem solving systems. In addition to the traits such as problem-solving and critical thinking, there are other cognitive attributes assigned to technology use such as attributes associated with flow (engagement and time dissociation) and attributes associated with computer control. Thus, there is little difference between virtual and real-world interactions. In particular, there is little cross-over of attributes between factors suggesting that each attribute is domain (factor) is specific and local independence. This also suggests that there are higher-level or more general attributes. These general attributes may be those attributes that are "upstream," from the current set of attributes. As the subjects deal with multiple set of information found in the SEG environment while solving complex problems, cognitive load (A5) becomes a consideration. However, review of the A5 in light of the hierarchical nature of the attributes suggests that as load levels increase, it (A5) begins to inhibit other factors. The ANN model assigns negative coefficients when they act as inhibitors of other factors. Initial game behavior seemed to focus on F1 as the subjects engaged in exploration and transitioned to F3 as load levels decreased.

CHAPTER THREE

Method

The validation of the pilot study suggests that the methodologies used within the pilot study are valid and usable for the larger study with some slight modifications to address questions regarding the design of SEGs and the underlying cognitive attributes. In particular, the use of factor analysis loadings, neural network weightings, and methods for the assignment of cognitive attributed to particular tasks. The following section of methods refers to methods used in this dissertation study specifically focuses on the design and not the play of SEGs.

Items as the Unit of Analysis

With the explosive growth of resources available when using computers, it is important to find ways to seek out and analyze useful information from these large amounts of user-generated data. Methods based in item response theory (IRT) and data mining provide for a useful framework from which to develop item parameters for these large data sets. This study used a combination of computer usage data, data mining techniques, and more traditional item analysis methods found within measurement, to capture, model, analyze behaviors, and task completion patterns exhibited by subjects. Within this framework, it is important to understand that the data derived from the computer usage logs is the unit of analysis. Computer usage log data-mining uses

secondary data retained in server logs, user profiles, registration data, user sessions, transactions, mouse clicks, and other data. More specifically the data consists of a unique identification number, gender, race, school level, state, Internet protocol address (IP address), and action identification number. The action identification numbers (task) are coded 1 through 45 and considered key tasks by the designers to complete the design of the SEGs. It is important to understand that the study author did not select the tasks and that these tasks are intrinsic to the design software. Table 1 describes the codes and the corresponding tasks.

Table 1

Data Code Description

Code	Task Description
1	Science Based Task Editing
2	Initiated Game Play Mode
3	Initiated Editing Mode
4	Completed Science Content Based Task
5	Play Science Based Task Within Game
6	End Game Play Mode
7	Complete Science Task Development
8	Add Session Notes
9	Add Objects to Game
10	Delete Objects from Game
11	Add Text to the Game
12	Science Quiz Question Answered Correctly
13	Quiz Questions Added
14	Quiz Questions Deleted
15	Decision Point Added
16	Decision Point Deleted
17	Total Quiz Questions Added
18	Edit Decision Point
19	Quiz Notes Added
20	Total Quiz Questions Deleted
21	Total Triggered Events
22	Total Decision Points
23	Successfully Completed Science Quiz
24	Delete Text From the Game
25	End Editing Mode
26	Successfully Completed Science Tasks
27	End of Level Achieved
28	Total Levels Edited
29	End of Game Achieved
30	Delete Session Notes
31	Total Time In Game
32	Map Code
33	Total Data Stream
34	Game Check Point Achieved
35	Quiz Choices Added
36	Quiz Saved
37	Decision Point Triggered
38	Total Decision Points Added
39	Total Decision Points Deleted
40	Decision Point Mapped
41	Total End of Games Initiated
42	Total Text Characters Added
43	Total Text Characters Deleted
44	Total Games Initiated
45	Total Session Notes Made

The data discussed within this study consists of weblogs of subjects' activities, coded while building SEGs designed around the generalized concept of science

processing. The discovered patterns via data mining in the item responses are representative of probabilities of task completion related to science content.

An important task in the computer usage mining is the creation of suitable pre-processed data sets. Meaning that the data must be cleaned and usable coding created from the larger conceptual codes created by the computer. The purpose of the pre-processed data is to offer a structurally reliable and integrated data source for pattern discovery. The total number of items used in the analysis n=154,240. In the application (the server logs) data was stored in extended log files, formatted as a comma delimited (CSV) file in excel.

In summary, data collection occurred via real-time, server data logging, as the subjects develop SEGs using a video games design system. Each subject's unique identifier assists in data sorting from the server logs. Server log data consists of all subjects' actions taken during the development and test play of the SEGs.

Sample

The unit of analysis for this study is the subjects' responses, i.e. mouse click and keystrokes. However, for the purposes of clarity, it may be illustrative to identify the human sample characteristics. The target population in the full study (not the pilot study) was subjects located in the mid-Atlantic region of the United States. Targeted subjects consist of students enrolled full-time traditional high school science program at grade 9-12 levels. Subjects' ages ranged from 14 to 18. Subjects within the study have taken a science class within the last semester (Fall 2012 or Spring 2013). Science classes considered in this study are Earth Science, Biology, Chemistry, or Physics. Criteria for

selecting subjects included; (1) taking their current science class for the first time; (2) taking the course as a member of a class and not in an online or virtual capacity; (3) admitted into the class within the first two weeks of class. This study derived the data set from a preexisting data set using the target population description as means to screen data points.

The study used a proportionate stratified sampling approach of science students to generate the computer log data. The sample size of each stratum was proportionate to the population size within the school district of interest. This particular sampling technique provides a higher statistical precision compared to simple random sampling and allows for a smaller sample size. In addition, this sampling technique increases the probability of inclusion of specific subgroups within the sample (Wallander, 2009). Stratum parameters for the stratified sample are grade, gender, and science class. Due to the sequential analysis of this study, selection of results within each phase results in aggregation of group results to reduce the number of dimensions for analysis. This population is of interest due to the increased perception that exposure to STEM rich environments increases the likelihood that subjects' select STEM discipline based majors in college and STEM careers after college (Lamb & Annetta, 2009; Lamb & Annetta, 2012; Lamb, Annetta, Meldrum & Vallett, 2011). This work is a result of The National Mathematics Advisory Panel and the National Science Foundation assessment of the United States standing in the STEM disciplines as in jeopardy, and will lose its status and place within the early 21st century without increases in outputs from the STEM pipeline (Annetta, 2008).

Subjects have had prior exposure to some educational game development within their science course work and therefore possessed cursory domain knowledge but lacked task specific knowledge. The subjects also took part in classroom activities such as lecture, content instruction and mentor-guided research reports in an effort to build an understanding of their topics and drive game design. It is important to understand the intent of the study is not to measure student gains in content knowledge, but to understand the cognitive attributes associated with the design of SEGs.

Informed Consent

The principal investigator obtained informed consent from the parent or guardian allowing their child to participate in a study associated with the National Science Foundation (NSF), [1]GRADUATE. To examine review board approval for the GRADUATE study please see appendix E. During data collection, members of the team obtained assent from all students prior to the initiation of each component of the study. The principal investigator sent a letter communicating procedures and expectations to the subjects (students) and their parents / guardians. The primary investigator presented each student and their parents / guardians with a printed copy of the informed consent for review along with ways in which to address concerns and questions. The parent / guardian and the primary investigator provided their signatures as means to express

[1]This material is based upon work supported by the National Science Foundation under Grant No. 1114499. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation."

consent. Study subjects participating as part of the study received compensation in the form of iTunes gift cards and entrance into a competition for a scholarship.

Confidentiality. Maintenance of confidentiality occurred in the following ways:

1. Study organizers collected informed consent and assent separately. In addition, the consent was not a part of the protocols and questionnaires.

2. The study outside evaluator developed an identification code to identify each subject anonymously and disassociated the code with subject personally identifiable information.

3. Subject informed consent forms were stored in a locked file cabinet separated from other data sets.

4. Only, the outside evaluator had access to the identifying information.

5. Subject codes and resultant data were stored in an encrypted folder on a password-protected computer.

6. Upon completion of the study, the outside evaluator destroyed the subject master list of code numbers.

Risk to subjects. The nature of the research and intervention was such that the risk for harm, discomfort, or damage to the subjects was very low. No questions or actions within the study touched upon emotionally or physically sensitive topics. A review of the research procedures suggests that subjects may experience fatigue from extended computer use. The subjects and study directors mitigated risk through self-paced resting measures. Study protocols reminded subjects that participation within the context of the study was voluntary and thus no adverse outcomes were associated with withdrawal from the study. Finally, subjects were encouraged to express concerns and questions though multiple modes of communication to include discussion boards, phone, and emails.

Design

Based upon the pilot study discussed earlier, the current study design was modified to improve upon the pilot study design, in particular, the use of a one-group observational design as a means to minimize intervention bias. The study design is a one-group, test-retest, and observation only, design. The use of an observation only design aids in the evaluation of the validity of the results. Specifically, intervention variables do not confound observed results. Lack of pretest creates positive and negative outcomes for internal validity of the measurements (game design actions and tasks). Positive outcomes would suggest that there are no threats to internal validity due to pretesting. Lack of pretesting reduces the overall threat to internal validity of the task items by reducing item familiarity and carry-over (Dimitrov, 2010). Methodologists suggest this design type for exploratory studies such as this one to help establish initial study questions and

hypotheses. Embedding the study intervention in the classroom creates a more true to life classroom experience; and the quality of collected data may be superior.

Task Presentation. The subjects took part in the NSF funded study GRADUATE. GRADUATE is a Serious Educational Game design process grounded in a problem-based learning model. GRADUATE provides subjects with the ability to design a realistic virtual environment involving scientific problem solving and research work in a scenario, task-based presentation. The overall intervention took place over the span of 8-months and integrated into the normal curricular environment (science class). Prior to contact with the subjects (students) the research team met with teachers to establish lesson plans around the topic of alternative energy use. The teachers worked with the Primary Investigator, staff, and mentor scientists to create lesson plans. Teachers taught lessons during normal instructional units and times to minimize interruption to the normal curricular flow. As part of the units study subjects conducted independent research on their topics related to alternative energy. In addition to the independent research, the subjects wrote an extensive research paper and met with mentor scientists who provided guidance and assistance on the topic. Based upon their research and meetings with the mentor scientists the subjects initiated the design of the SEG with the creation of a storyboard. Subjects presented their storyboards to peers, their teacher, and the mentor scientists for feedback and review. Subjects integrated the feedback and review into the design of the game along with the modified storyboard in an iterative process. Upon completion of the storyboard and reports, students initiated the actual use of the SEG design software. During the design of the SEGs, students used classroom discussion,

chats, feedback, and design testing as a means to modify and complete their games. Figure 1, illustrates a graphical view of the overall game design process. Note that while there seems to be a singular linear pathway for the overall design process, subjects actual sequences through the design process were not always linear. Each component of the design process provides a framework for the work of the GRADUATE project. For example, conducting research for the research paper would occur under the Gather Information portion of the framework. Infusing pedagogy a key component of the SEGs game design process occurs within the Analyze Information portion of the framework. Prototyping, testing and improvement would occur within the Testing and Feedback, Improvement and Develop Products portion of the framework.

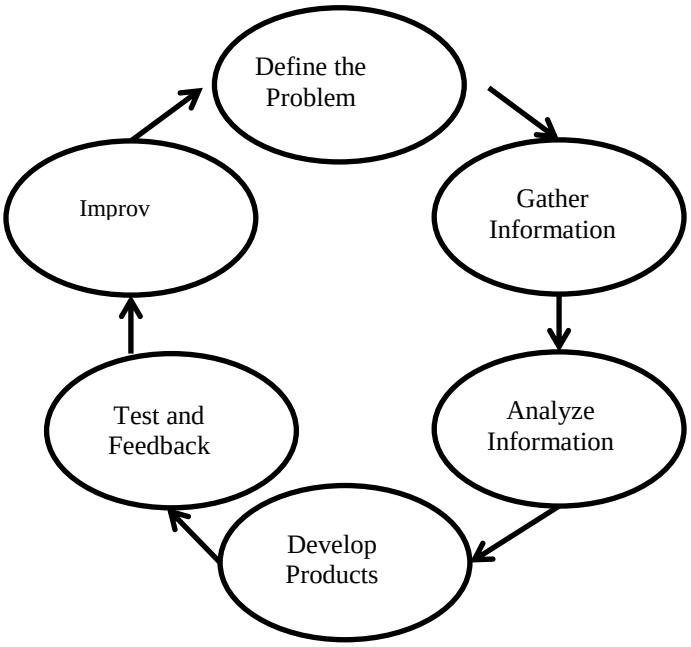


Figure 1. Iterative Design Process for GRADUATE**.**

Subjects designed SEGs focused on concepts related to alternative energy as presented across the five principal domains of science (Earth and Space, Biological Science, Chemistry, Physics and Environmental Science). Upon initial login, subjects receive a description of a learning problem to solve using a games based design approach. As the building and design of the game progresses, the subjects (game designers) assign characters to complete tasks in keeping with game completion and ultimately this supports learning. The subject is then required to design various science-based, problem-solving tasks in a SEG environment to complete their video game building. The complex tasks involve, but are not limited to, storyboarding, creation of a learning scenario, and design and placement of objects in a three-dimensional working environment. Through these processes, the subject identifies key learning tasks through experimentation, teacher led instruction, readings, and singular task-based training.

The subjects of the study also designed and answered items in a drag and drop interface (matching). Completion of the particular task items result in their progression through the game design process. The game logs each action (mouse clicks, inventory interactions and questions answers, etc.) taken in the game, on a per session, per player, basis. During data cleaning, and coding of subjects' task completion outcomes used a dichotomous approach. Data were coded either a "1" for successful completion of the task item, or a "0" for unsuccessful completion of the item. Through completion of the complex tasks, subjects are able to show mastery of simple tasks, subtasks, and content knowledge related to the games design and underlying science processes.

Based upon the identification of the factors and tasks, the researcher assigned tentative cognitive attributes found in the literature to the tasks. The cognitive attributes identify the underlying components of cognition needed to complete tasks. After assignment of the cognitive attributes to the tasks, the study author presented the resulting matrix to expert reviewers for an assessment of relevance for each attribute task combinations. Reviewers designated the assignment of the attribute to the task as *Strongly Relevant* (SR) or *Weakly Relevant* (WR). The assignment of the relevance rating is based upon, the supplied supporting literature, experience and expertise, and nature of the task-attribute relationship. These results develop into a matrix outlining the relevancy combinations.

Analysis

This study intertwines the two disciplines of data mining and statistics as a means to analyze the large data streams. To understand the role each plays in the development of the data structures, a brief history and outline of statistics versus data mining is necessary. Statistics is the mathematical science pertaining to the collection, analysis and interpretation of data (Steen, 2010). Data mining is the process of collecting, analyzing, and identifying patterns found in large data sets (Ngai, Hu, Wong, Chen, & Sun, 2011). Both statistics and data mining have common goals. One primary goal of both is to understand the structure of the data. However, in addition to this goal, the aim of data mining is to make use of computational methodologies such as artificial neural network analysis (ANN) to develop predictive pattern recognition. This goal (pattern recognition)

makes data mining a suitable methodological approach for CDA. However, one specific difference between data mining and statistics is that many of the data mining methodologies make use of the ad hoc analyses, coupled with data driven models, as opposed to those models derived from theoretical approaches such as those found in statistics. While the differences between data mining and statistics may not be readily apparent, historical context provides a means to justify the use of data mining or pattern seeking within large data sets. The mathematical background and history of statistics encourages a tendency to require evidence that a particular methodology is successful prior to its employment. One can necessarily contrast statistics in many ways with the history of data mining, which, arises from computer science and machine learning research methodologies. In practice, this contrast with data mining (lack of evidence), results in the use of methodologies, that provide insight into the nature of the learning without a foundation based in proven outcomes. Given that the development of statistics predated the invention of the computer and more recently the invention of the parallel processing in the mid-1990s, many of the statistical techniques were developed employing "hand" calculation methods; utilizing samples smaller than 1000. While rigorous in nature, the older techniques (generation one) were often unable to handle the large data sets now generated by computer-enhanced learning (i.e. computer data logging). The principal outcome and problem with large data sets in statistics is the development of statistical significance due to minor variations that are not practically significant due to the large sample size. Thus, it becomes a problem of practical significance versus statistical significance. To place this differential into perspective, the

typical statically manipulated data sets of 10,000 data points is 13,000 times smaller than the data points generated via computer data logging on a daily basis for the Visa Corporation (retrieved: 3/10/2013 http://corporate.visa.com/about-visa/technology/transaction-processing.shtml) when analyzing purchase data for its customers. In regards to this generation, one statistical analysis would be woefully inadequate to develop understanding from this vast data set. From this difference, it is clear that manipulations of data at this scale require the use of a computer and methodology with the capacity for pattern searching such as data mining.

Analysis of this study's large data sets requires a combination of data mining techniques and inferential statistics developed in a measurement framework such as IRT and cognitive diagnostics. The study analyzed the data in three interconnected phases: phase 1, dimensional reduction using exploratory factor analysis cross-validated with a confirmatory factor analysis, phase 2 psychometric analysis of task responses (subject movement through the game design) using a 2PLM IRT model for parameterization, and phase 3, development of the of the Q-matrix and development of the artificial neural network (ANN).

Phase 1 consists of randomly dividing the sample in half, while maintaining strata. The study then compared the two subsamples to ensure equality of sample and variance. Exploratory factor analysis (EFA) outcomes for ½ N to assist in dimensional reduction and identification of the unknown factors that underlay the direct measures associated with game construction related to science content. Three experts in the field of science education and educational psychology examined the results of this portion of the

analysis aiding in identification of relevance of the cognitive attributes and perceived validity of their association with tasks. The cross-validation of the EFA results used the second half of the sample through a Confirmatory Factor Analysis as suggested by Dimitrov (2011). The resulting model provides the framework from which to create the task item parameters results for the second phase of the study.

Phase 2 is a psychometric analysis of subject responses to the task sets. This phase involved the use of item response theory to validate the assessment, model fit, item constructs, item functioning, along with probabilities. In this case, the assessment is the successful completion of tasks associated with the build of the subject-developed video games (SEGs). Parameterization of the probabilities occurred by calculating the odds of task completion, using a "1" for success and "0" for failure to complete the task. The log-odds of completion provided the information pertaining to the probability of successfully completing both content tasks and game design tasks. Specifically, the parameterization encompasses a comparison of expected completion, versus actual task completion, measured through X^2. Item (task) fit analysis is conducted using a two-parameter model (2PLM) (*a* and *b*) with resulting infit and outfit statistics providing model fit information.

The results of phase 1 and 2 inform phase 3, and the development of the Attribute Mastery Pattern (AMP) in the form of a Q-Matrix and artificial neural network. Development of the AMP involved identification and validation of the cognitive constructs that underlie the task items via expert review. These phases result in the creation of expected response patterns and attribute probabilities. The study presented the

results from this phase to an artificial neural network in the form of training and test data to establish attribute hierarchy via propagation weightings and model fit.

Summary of techniques. This study used several statistical and data mining approaches in order to develop a repeatable, methodological approach to the measurement of cognitive attributes in Serious Educational Game design. The methods provide information for each of the follow-on analysis. Figure 2 provides an overview of the methods used to answer each research question.

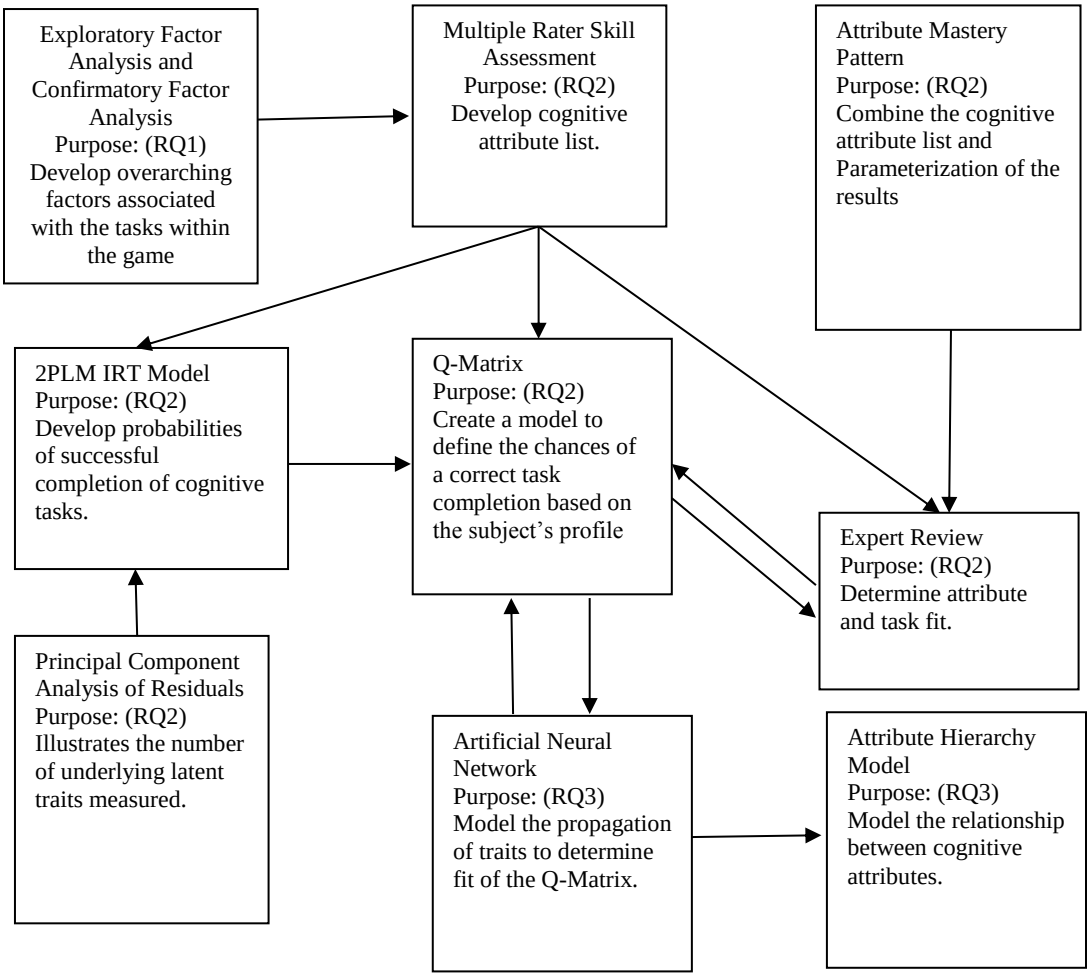


Figure 2. Summary of Methods and Related to Each Research Question

Exploratory factor analysis. Results of the exploratory factor analysis and confirmatory factor analysis, answered research question 1, *what are the underlying factors exhibited through the measurement of task items associated with subject development of Serious Educational Games?* Initial exploration of the data structures methodologies indicate the use of exploratory factor analysis (EFA), cross-validated with confirmatory factor analysis (CFA), to simplify, explain, and confirm the complex variables and the relationships among them. EFA provides a means for uncovering underlying clustering patterns within the data when there is little underlying theoretical framework. Construction of the factors that underlie the relationship between item results from EFA clustering. More specifically, we can determine how many factors underlie a set of variables and which variables form which factor. Construction of the factors, which underlie the relationship between the items results from the EFA loading. More specifically, one can determine how many factors underlie a set of variables and which variables form each factor. Validation of the EFA occurred via the use of confirmatory factor analysis. Through maximization of the total variance, one can choose the most appropriate factors for the observed variables.

Using Varimax rotated EFA; the resulting structure provides the greatest variance in accounting and produces a simple solution with one major loading per factor. The study used the simple solution to develop an understanding of the relationship between items and factors, along with the proportion of the variance accounted for between the

factors as the partial loading. Each of the factors represents the maximization of the total variance in the observed variables. Methodologists suggest the use of exploratory factor analysis (EFA) when there is insufficient theoretical information to hypothesize the number of underlying factors accounted for by the variables. As this study is an exploratory study, there is little theoretical foundation to draw from.

Researchers use the principle factor method (PCA and ICA) for analysis of the underlying factors to confirm and explain the underlying latent constructs of a measure. Extraction of the first factor is contingent upon maximization of the variances accounted for in the primary factor. The process then removes the factor and the next factor, which accounts the maximum variance, is calculated. Analysis continues through this process until it accounts for all variance in the resulting factors. Low correlation between the items structures can suggests a trend toward an orthogonal factor structure maximizing the variance across all factors. If the resulting factor model produces orthogonal factors, which when rotated, produce more interpretable results. The resulting factors from the CFA confirm the attributes related to the factors. The research validated the correlation between factors using Pearson's *r*.

Item response theory (IRT). Results of the IRT analysis assisted in answering research question 2, what are the cognitive attributes that underlie the design of Serious Educational Games? Researchers use IRT analysis primarily for the development, evaluation, and validation of assessment instruments. Instruments developed using IRT analysis contains items or tasks, which remain fixed along a scale allowing for calibration across differing samples. The use of two-parameter measurement models constructed under the paradigm of IRT provides a theoretical model to create an equal measure analysis of the embedded assessed items, which equated with tasks in this study.

The two-parameter IRT model is probabilistic and based upon the logit function (Rasch, 1960; Linacre, 1991 & 1999; Chen, Wong, Leung, & Kwan, 2012). This probabilistic model allows for an adequate measure of those items (tasks) which the subjects are least likely to be successfully complete. Individuals who exhibit a higher likelihood of exhibiting a greater completion level are more likely to show increases in task completion. Consequently, when a high measuring subject does not complete items that are ranked lower, those endorsements are considered unexpected and result in larger outfit deviations. A second advantage of the two-parameter IRT model is that the model provides for the construction of a linear measure through transformation from ordinal observation and quantification of the response categories within the task list (Linacre, 1999). The construction of the linear measure, from the ordinal data set, derives from the transformation of raw scores, to the common metric of logits (inverse of the sigmoidal function). The transformation of the measure responses and item calibrations occurs via comparison of the data to an existing IRT model. The ordering of the items on the

response measure creates an additive relationship allowing for the development of probabilistic models (Betemps, Smith, & Baker, 2004). Probabilistic models allow for statistical comparisons of the expected responses to the actual responses within the model using χ^2. From the comparison of expected task completions to the observed task completion, it is possible to provide an indication of IRT model fit. The use of the single parameter model is only applicable to the characterization of single trait constructs such as that of specific cognitive attributes.

IRT and the Q-matrix. The resulting parameters *a* and *b* link the IRT model to the Q-matrix via marginal true-score measures for binary items (Dimitrov, 2003). Researchers use the Q-matrix to calibrate the probability of task completion. Probabilities derive from using IRT develop the parameters of the Q-matrix. Within the Q-matrix, the inverse linear-equal measure interval probability is a measure of the extent to which abilities not specified in the developed Q-matrix, affects the probability of correctly completing a task. The task completion probability is the inverse to the difficulty parameter in the two-parameter IRT model. Lower values imply an influence from abilities not specified in the Q-matrix. Rasch in 1960 proposed an equation that is the derivation of this model for cognitive validation. Based upon combination of the Suppes's probabilistic model, (Suppes, 1969), and the Spada probabilistic model (Spada, 1977). Dimitrov (2007) combined these two models to account for the differences in error rates among individual subjects creating a uniform step model for clearer parameterization of the Q-matrix.

Attribute mastery pattern (AMP). The probabilities developed using the IRT model integrates into the attribute mastery pattern via the Q-matrix. Development of the attribute mastery pattern uses only dichotomous values, indicating either the presence or absence of the cognitive attribute related to the proposed task. Using a componential approach, such as Attribute Mastery Patterns (AMPs) applied to factor analysis as a means, reduces the factors to a minimum number of dimensions. The researchers then determine the number of underlying components of the AMP from the reduced factors (Sternberg, 1982). Three experts in the field of science education and psychology made the determination of presence of the attribute within the construct through review of factor-item relationships. Suitable agreement (agreement coefficients of 0.70) between reviewers indicated the presence of the cognitive attribute (Lamb, Annetta, Meldrum, & Vallett, 2011). Further analysis and confirmation of the proposed cognitive attributes via verbal report studies will support selection of this model.

The Attribute Mastery Pattern (AMP) is a method of item response classification that categorizes subject responses based upon cognitive models of task performance (Leighton Et al., 2004; Leighton, Gierl, & Hunka, 2006). Task performances are representative of the underlying cognitive traits and attributes. Items within the attribute mastery pattern are related in a nonlinear-network using Bayesian probabilities, the approximation of which is possible with artificial neural networks (ANNs). This non-linear network represents the interrelationship between the attribute competencies, the measurable factors, and the items (Anderson, Douglass, & Qin, 2005; Kuhn & Matson, 2002). An Artificial Neural Network uses in this capacity validated model fit, i.e. to test

the AMP. Specifically, the cognitive attributes within the AMP network represent the procedural knowledge and processing functions needed to perform particular domain based tasks (Wang, Jackson, & Zhang, 2011). The IRT model extracted the subjects' probability of mastery in relation to factors using neural networks parameterized weightings. The presentation of response patterns to the neural network determined model fit, paired with comparison of the results to the expected outcome. Calculation of model fit for the AMP uses the following equation, Equation 2 proposed by Cui, & Leighton (2006):

$$HCl_i = 1\text{-}2\sum_{j=1}^{J} \sum Xi_j(1\text{-}Xi_g) / N_c \qquad (2)$$

Wherein *J* is the total number of items, X, is the subjects score to item j, S, includes items that require the subset of attributes of item j, and N_c is the total number of comparisons for the correct answers. Values for the AMP are between -1.00 and +1.00. Mean values above 0.60 indicate adequate model fit while values below 0.30 suggest poor model fit (Wang & Gierl, 2007).

Artificial neural networks (ANN). An artificial neural network (ANN) is a method of computing relationships based upon the interaction of multiple, connected, processing elements in a non-linear fashion (Pinkus, 1999; Gupta, 2010). A key feature of artificial neural networks is that there is a strong connection between input elements and output elements. However, the elements dealing with the input-output relationships are not fully known or researchers could model these connections directly. Artificial neural networks represent a new paradigm for the analysis of complex human emotional and

cognitive constructs, such as the construct of interest and items relating to cognitive attributes. Designers of ANN models base them upon an abstraction of higher-level, cognitive functions associated with neural "wetware" (i.e. vertebrate brains). Specifically, the ANN and derived statistical models mimic the architecture of the parallel, non-linear processing found in organic based brain systems. Represented statically in the form of a graphic is the parallel architect that provides an understanding of emergent relationships (patterns such as those found in data mining). These ANNs are often used in three modes: (1) as a model of biological nervous systems and intelligence, (2) real-time, adaptive, signal processors, and (3) as a data analytical methods. This study uses ANNs in the second and third capacity as a real-time adaptive signal processor of cognitive inputs, and a data analytical method to theorize and model task completions. The generalization of this artificial neural network model reduces to the underlying functions, algorithms, of pattern recognition. Given this understanding, it is important to remember that researchers represent the patterns of an ANN in the terms of numerical values assigned to nodes within the model. The numerical values transmitted along the network use an algorithm for propagation. It is important at this point to differentiate between the ANN models and ANN algorithms. One of the major differences between ANN models and ANN algorithms is the manner in which data is used. The ANN algorithms are more appropriate to analyze transient data thus making them relatively useless as a statistical test procedure. However, ANN models are far more useful for developing the repetitive analysis necessitated by statistical algorithms. While there are considerable similarities

between the statistical models and ANN models, there are differences within the terminology and language of ANNs.

Developers of Artificial Neural Networks designate nodes using one of three descriptions. The designations are input nodes, output nodes and hidden nodes. The nodes link using weighting parameters, thus the input nodes, hidden nodes and output nodes become a multivariate function similar to the concept of a non-linear structural equation model (Dijkstra & Henseler, 2011). This nonlinear function describes the movement and transformation of task processes, via weightings along the network nodes. The ANN accomplished the actual transformation of the parameter estimates via learning algorithms that include the use of backward propagation in the case of multi-layer perception delta-rule networks such as the ANN in this study. This propagation makes the network model adaptable by adjusting the weighting by a proportional difference between the expected output and the actual output. In addition to weighting adjustments, it is possible to standardize the output of the maximum propagation weight to 1.00 as in this study. This adaptive ability allows for flexibility within this model not seen in other modeling techniques. Smith and Gupta (2003) proposed equation 3, the general equation for adaptation and propagation within an ANN:

$$\Delta Wi = \acute{\eta} * (D - Y)Ii \qquad (3)$$

Where $\acute{\eta}$ is the learning rate of the ANN, D is the desired output and Y is the actual output. The above equation is an extension of the Perceptive Learning Rule (also

known as the Hebb's Synapse and learning rule) (Hebb, 1949). Figure 2 provides a proposed model of an artificial neuron and its propagation to output.

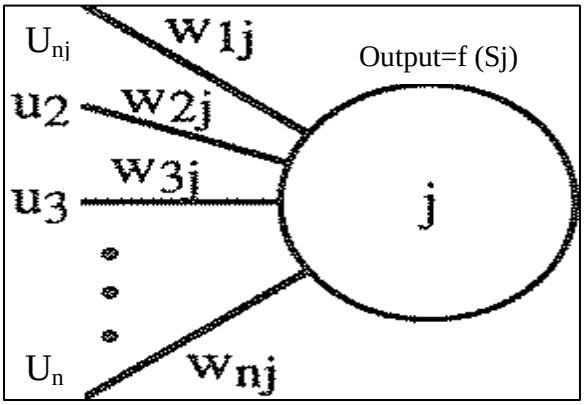


Figure 3. Artificial Neuron, the Functional Portion of the Neural Network

The movement from a narrow view of ANN as an adaptive message-passing algorithm to a statistical processor involves the inclusion of probabilistic assumption regarding the data in particular the input nodes and output nodes. The development of probabilistic assumptions for the ANN in this study derives from the 2PLM IRT parameters. The ANN represents the input nodes as patterns, which appear in the input, while the output nodes are recast as resulting samples of a density of higher–dimension, randomized, probability estimations. It is this link to inferential statistics, which allows linkage of an ANN to practical descriptions of real-world problems such as cognitive-attribute, task completion probabilities, and development of the hierarchical relationships. It is in this light that the ANN offers answers to a complex array of problems though its

intricate statistical modeling with an emphasis on flexibility. However, this inherent flexibility is sometimes the cause of overfit errors, which increases as the increase in variables creates randomization of components resulting in a decreased performance for future data. To control for this, it is important that the data have a similar level of, or greater level of, complexity than the existent data. The use of a back propagation, regularization and other Bayesian methodologies results in a decrease in the number and type of overfit errors.

Statisticians classify the statistical approach to ANN in the class of statistical models for nonparametric statistics, thus, are not subject to assumptions of normality and sample size. Some examples of similar statistical models are the Generalized Linear Model, Maximum Redundancy Model, Projection Pursuit, Cluster Analysis, and Radial Basis Function (Orr, 1995). If the model does not contain hidden layers and instead maintains direct connection between the inputs and output neurons, the model becomes a Functional Link Network. This Functional Link Network is akin to the statistical term known as main effects analysis. The generalized equation for neural network propagation shown in equation 4:

$$x_n^i = F(y_n^i) = F(\sum_{\ell=0}^{C_{n-1}} w_n^{i\ell} \cdot x_{n-1}^{\ell}) \tag{4}$$

Methodologists suggest the use of a stepwise neural network to ascertain the number of cognitive attributes and their hierarchical relationship contained within the proposed model. Fit statistics found in the pilot study suggest that a model using 16

attributes is the most parsimonious model. Using the statistical software package JMP 10.0, the results (item responses) were presented to the artificial neural network.

The nodes within this model fulfill different functions. The input nodes within this network present data to the remainder of the network, with each node containing one piece of the larger data items. Within in this model, a data item is one task. Each of the hidden nodes indicates the presence or absence of a particular attribute. The model then has two output nodes, each node representing either successful completion, or unsuccessful completion of the task. The output nodes also represent an additional factor of strength of propagation allowing for the development of hierarchal relationships between each of the attributes. During this study, resultant ANN, weightings normalized the outputs to sum to 1.00 (100%). Thus, the node with the highest value is basal attribute for the remaining two subordinate attributes.

Further analysis of the ANN outputs occur using Independent Component Analysis (ICA). ICA is a clustering algorithm based in neural outputs and found in JUMP 10.0 helping to assign grouping of cognitive attributes to tasks within the Q-matrix. ICA can provide outputs where such traditional analyses such as Principle Component Analysis have failed (Beckmann, 2012; Bingham, Kuusisto & Lagus 2002; Chang, 2012). Clustering of this nature confirms prior analyses of factors relating to attributes developed during phase 1 and 3 of the analysis.

ICA hidden layer analysis. ICA differs from PCA in a key assumption relating to the treatment of the data post decorrelation and variance accounting. ICA requires an additional data transformation to develop independent components. This assumption

plays a particularly necessary role when dealing with non-normal (non-Gaussian) distributions such as those found at the school or classroom levels of educational data analysis. A violation of the normality assumption is of concern because when using a non-normal distribution, independence and non-correlation differ with independence being the stronger property (Lechner, Lollivier & Magnac, 2008). When using ICA, one assumes that independent components lack a Gaussian distribution as opposed to PAC in which the analyst assumes components are normally distributed. Taking this distribution assumption into account then, ICA is equivalent to PCA. However, if the data is not normally distributed and multi-dimensional ICA is superior as it (ICA) can take complicated multi-dimensional data and identify its underlying structure.

Selection and design of training and test data. Science processing is a complex domain; researchers must take care to ensure that modeled tasks are not overly complex for this exploratory study. It is not computationally possible to model all process attribute types. The goal is to select tasks that would reflect the domain complexity and provide generalizable results. Limiting of task selection occurs through using tasks already intrinsic to the game software design-process. In particular, the 45 tasks identified are those tasks. Therefore, to successfully generate a model and balance the computational concerns with useable tasks, considerable care in task selection is required. The study models allow duplication of task types within the science processing domains as task types often overlap between multiple tasks. During this study, a panel of experts validated core-task selection, within the domains to ensure applicability. Each task was assigned to a factor and randomly assigned to the training set or test set. This ensured that

there is no crossover between testing sets and training sets. Encoding of the attributes parameters occurred via the development of the Q-matrix.

Combining Bayesian models and IRT models. Parameterization of the task completion likelihoods and the use of ANN training models (Bayesian Approaches), as in this study, assist in the development of more effective targeting of tasks to attributes. The use of these particular models helps to develop an effective picture (model) of individual subject cognitive processes for simulation and teaching purposes. Effective ANN development provides researches with an effective means to simulate and test learning. To create input vectors for each of the science process tasks it is necessary to encode probabilities of successful completion of the task item. Transformation of initial responses occurred using an IRT model, specifically the 2PLM model. The study used two-parameter logistic model parameters to compute the population probability using the IRT True-Score method. Based upon the results of the IRT True-Score tasks probabilities for the population, individual probabilities are assigned to cognitive attributes using a Q-matrix and the artificial neural network propagation weightings Node coding developed using one input node per attribute; flagging of the node via a "0" or "1" indicates the presence or absence of the attribute. This type of coding provides a simpler model allowing the ANN to learn the input parameters more efficiently (Bishop, 1995; Bhatt, 2012). Folding all values of the parameters into one node and all constants into another node is a way to represent and account for prior knowledge. The accounting for prior knowledge within the ANN model is a key feature of Bayesian models, which are not present in IRT models (Soares, 2009). Since propagation across the network is contingent

upon the presence or absence of the attribute. The attribute values used to determine success are not of consequence to the solution and its propagation across the ANN. Coding input vectors (tasks and attributes) in this manner permits the coding of a large number of examples of science process tasks which preserve individual identities of the tasks. Due to the potential for a larger number of coded tasks, the ANN model becomes more flexible and generalizable as the number of parameters increases. This also reduces the likelihood that review bias concerning relative importance of one attribute versus the others affected the results.

A second area of strength, thus indicating the mixing of IRT and Bayesian models as superior, is in the number of attributes required for analysis. Model convergence using an ANN occurs with smaller number of attribute to item-task ratio. Researchers using this modified form of cognitive diagnostics can obtain usable result with fewer suggested attributes resulting in easier interpretation for practitioners. Fewer attributes make it more likely that the educator can successfully target those attributes during instruction in a timely manner (Huff& Goodman, 2007). The use of fewer attributes in-turn can help to increase interpretability of the data at the classroom level (Roberts & Gierl, 2010). Through a combination of the Bayesian and IRT models researchers are able to capitalize on strengths of each while accounting for weaknesses.

CHAPTER FOUR

Results

The results section is organized by analysis type and research question. The three research questions answered are; RQ1, *what are the underlying factors exhibited through the measurement of task items associated with student development of Serious Educational Games?* RQ 2, *what are the cognitive attributes that underlie the design of Serious Educational Games?* RQ3, *what theoretical mathematical / statistical model develops using an Artificial Neural Network to describe the interaction of the items, factors, and attributes as subjects design Serious Educational Games?* The main results of this study include the theoretical connection between the games and attributes through the Q-matrix. A second result is a model of Serious Games as assessment. The third result is the identification of tasks and attributes used in the design of SEGs, and lastly, an artificial neural network used as a model to investigate how subjects learn science while designing Serious Education Games.

Exploratory Factor Analysis

Results of the exploratory and confirmatory factor analysis inform RQ1; *what are the underlying factors exhibited through the measurement of task items associated with student development of Serious Educational Games?* Results within this section suggest rejection of the null hypothesis (Ho $\Lambda i=0$). The Kaiser-Meyer-Olkin measure of sampling

adequacy was 0.764 and the Bartlett test of sphericity was significant (χ^2=3579.78, *df*=780, *p*<.001). Each of these statistical results suggests that the data is appropriate for factor analysis. Initial inspection of the scree plot (Figure 4) of eigenvalues evidences a departure from linearity coinciding with a 4-factor solution.

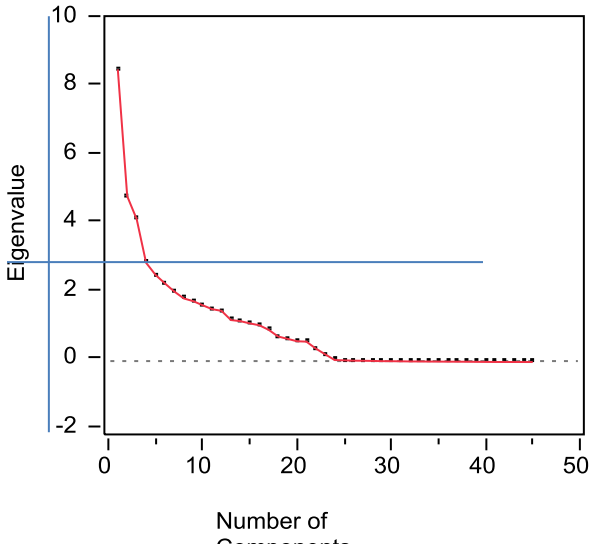


Figure 4. Scree Plot of factors.

The EFA loading coefficients were examined for their resemblance to a simple factor structure. Meaning that in an ideal case, each item would have one large loading on a single factor with all others loading close to zero. Methodologists suggest the removal of items that do not illustrate a simple structure. Removal of these items from the model aids in the development of a clean construct. A second consideration is the cutoff value for loadings. Portney & Watkins (2000) suggest a cutting score of 0.30 as a sufficient

loading (Portney & Watkins, 2000). Consideration of these criteria resulted in the removal of 11 items for low or improper loadings. The rotated factor matrix shown in Table 2 suggests that four factors account for 80.75% of the total observed variance in the measure. Analysis of the factor loading items suggests the factor loading descriptions. Factor 1 consists of items; 1,4,5,7,12,23,26,27,29, and 34. Each of these items (1,4,5,7,12,23,26,27,29, and 34) concern subject outcomes to science process skills tasks. Each task relies on a subjects' understanding of science process for success. From this relationship it is suggests that the factor be labeled Science Process. Factor 2 consists of items: 2, 3, 6, 8, 9,10, 11, 24, 25, 28, 30, 31, 32 and 33. Each of these items suggests a factor related to game editing and control tasks. An example of this type of item is the item (task) 9, Add Objects to Game. Results suggest Factor 2 is Game Control. Factor 3 consists of 5 items; 13,14,19,35, and 36. Each item suggests the addition of quiz item(s) within the game related to science based task items. An example of this type of item is item (task) 13 Quiz Question Added. A label, which may be applied to this factor, is Quiz Development. The final factor, Factor 4, consists of five items; 15, 16, 18, 37 and 40. Each of these items is associated with the development of decision points within science game contexts. Examples of these items are Edit Game Decision Points. An appropriate label for this factor is Game Logic. Relationships between the items suggest the label of Game Decision Development for this factor. Results suggest the removal of several task items from the model due to factor loadings below 0.30. Confirmatory factor analysis aided in the development of greater clarity. This clarity arose from examination of the factor structure established using the confirmatory factor analysis (CFA).

Table 2

Task Factor Loading

Task Number	Task Description	Factor 1	Factor 2	Factor 3	Factor 4
1	Science Based Task Editing	0.91			
2	Initiated Game Play Mode		0.88		
3	Initiated Editing Mode		0.60		
4	Completed Science Content Based Task	0.48			
5	Play Science Based Task Within Game	0.48			
6	End Game Play Mode		0.50		
7	Complete Science Task Development	0.54			
8	Add Session Notes		0.40		
9	Add Objects to Game		0.38		
10	Delete Objects from Game		0.35		
11	Add Text to the Game		0.30		
12	Science Quiz Question Answered Correctly	0.63			
13	Quiz Questions Added			0.91	
14	Quiz Questions Deleted			0.31	
15	Decision Point Added				0.30
16	Decision Point Deleted				0.37
17	Total Quiz Questions Added	Suggested for Removal, Did Not Load			
18	Edit Decision Point				0.93
19	Quiz Notes Added			0.27	
20	Total Quiz Questions Deleted	Suggested for Removal, Did Not Load			
21	Total Triggered Events	Suggested for Removal, Did Not Load			
22	Total Decision Points	Suggested for Removal, Did Not Load			
23	Successfully Completed Science Quiz	0.94			
24	Delete Text From the Game		0.87		
25	End Editing Mode		0.59		
26	Successfully Completed Science Tasks	0.51			
27	End of Level Achieved	0.47			
28	Total Levels Edited		0.50		
29	End of Game Achieved	0.57			
30	Delete Session Notes		0.39		
31	Total Time In Game		0.35		
32	Map Code		0.33		
33	Total Data Stream		0.30		
34	Game Check Point Achieved	0.62			
35	Quiz Choices Added			0.91	
36	Quiz Saved			0.30	
37	Decision Point Triggered				0.30
38	Total Decision Points Added	Suggested for Removal, Did Not Load			
39	Total Decision Points Deleted	Suggested for Removal, Did Not Load			
40	Decision Point Mapped				0.94
41	Total End of Games Initiated	Suggested for Removal, Did Not Load			
42	Total Text Characters Added	Suggested for Removal, Did Not Load			
43	Total Text Characters Deleted	Suggested for Removal, Did Not Load			
44	Total Games Initiated	Suggested for Removal, Did Not Load			
45	Total Session Notes Made	Suggested for Removal, Did Not Load			
Total Items		10	14	5	5
Cumulative Percentage		28.95%	28.06%	14.07%	9.67%

Confirmatory Factor Analysis

Examination of a confirmatory factor analysis on 1/2n confirmed the data structure and resultant model for task groupings. Estimation of the model parameters occurred using the maximum likelihood method. A series of four models were tested. Reviewed for maximum model fit are the sequence of modeling outcomes and summary statistics. As reported for comparison with those of the EFA in Table 3, are the factor loadings for the final CFA model in Appendix D. Within the framework for confirmatory factor analysis, one specifies the factor structure that is hypothesizes from the exploratory factor analysis, in this case 4-factors. Results of the CFA suggests an adequate mode fit for a uncorrelated four factor model using an imposed restriction of all factor correlations at zero, (WRMR=0.56). The use of the imposed restriction is indicated when the correlation between all factors is not statistically significant. The indications of unidimensionality allow for the use of IRT on each factor grouping and not Multidimensional Item Response Theory (MIRT). Model fit is adequate despite the significant chi-square as the chi-square statistics is sensitive to sample size. Review of modification indices suggests that the removal of the fifth factor (MI=12.00). The hypothesized 4-factor model illustrated adequate model fit (χ^2 =2.39, *df*=44, *p*>0.001, CFI=0.95, TLI=0.95 RMSEA=0.04, 90% CI RMSEA=0.01, 0.05). Inspection of the factor loading coefficients from the confirmatory model revealed that, as with the exploratory model, 11-task items did not load sufficiently on any of the four suggested factors. The remaining items as suggested in the exploratory factor analysis maintained the factor structure developed in the exploratory phase. Appendix B illustrates the

confirmed factor task- item structure. Mplus code for the confirmatory factor analysis is in Appendix A.

Table 3 displays the resulting tasks as a function of the suggested factors. The table also combines the results of the 2PLM IRT model analysis. The suggested task items indicated for removal due to poor parameterization are task items 9, 11, and 31-33. The total number of task items loaded on each factor range from 5 to 10. The factor with greatest number of task items is Factor 1. This is expected, as Factor 1 (Science Process) deals with the most complex of the tasks.

Table 3

Task by Factor Breakdown

Factor 1	Factor 2	Factor 3	Factor 4	Removed Due to Factor Loading	Removed Due to Lack of 2PLM IRT Fit
1	2	13	15	17	9
4	3	14	16	20	11
5	6	19	18	21	31
7	8	35	37	22	32
12	10	36	40	38	33
23	24			39	
26	25			41	
27	28			42	
29	30			43	
34				44	
				45	
10	9	5	5	11	5 Total Number

This study uses the results of the IRT analysis to answer research question 2; *what are the cognitive attributes that underlie the design of Serious Educational Games?*

Results of this section suggest rejection of the null hypothesis (Ho P_{Ai} =0). Results illustrated in Table 4 are for the population parameters for the tested tasks. Table 4 displays the population proportion of correct response on item *i* (π), the population estimate of the item error variance σ2 (*e*i), the population estimate of the item true variance σ2 (τi), and the population estimate of item reliability ρii. Review of Table 5 provides the overall descriptive statistics for the combined test tasks as P_i = 0.366, VAR (e_i) = σ2e =4.21, VAR (τi) = σ2 τ =51.05, RO_{xx}= ρ_{xx} =0.95. These results suggest the test population reliability parameter is high (ρ_{xx} =0.95). However, the overall difficulty of the test is moderate with 36.6% of the population correctly completing all tasks. Of the total items included in the final analysis, item 19 is the most difficult while item 6 is the easiest task to complete. Items, showing difficultly over +/- 2 on Table 5, were removed due to poor 2PLM model fit. Item 1 is the most reliable at ρii = .64 while item 14 is the least reliable ρ_{ii} = .10.

Table 4

2PLM Item Response Model (Parameters a and b)

Task Number	Task Description	Discrimination (*a*)	Difficulty (*b*)	Notes
1	Science Based Task Editing	2.56	1.00	
2	Initiated Game Play Mode	0.89	1.21	
3	Initiated Editing Mode	1.46	1.36	
4	Completed Science Content Based Task	1.61	0.72	
5	Play Science Based Task Within Game	0.95	1.02	
6	End Game Play Mode	0.72	0.87	
7	Complete Science Task Development	1.40	1.31	
8	Add Session Notes	1.09	2.58	
9	Add Objects to Game	0.58	2.08	Removed (a)
10	Delete Objects from Game	0.54	3.14	
11	Add Text to the Game	0.63	3.77	Removed (b)
12	Science Quiz Question Answered Correctly	4.48	0.85	
13	Quiz Questions Added	0.72	0.98	
14	Quiz Questions Deleted	0.83	1.95	
15	Decision Point Added	1.30	0.95	
16	Decision Point Deleted	0.99	1.18	
17	Total Quiz Questions Added	0.66	0.84	Removed EFA
18	Edit Decision Point	1.54	1.24	
19	Quiz Notes Added	1.17	2.24	
20	Total Quiz Questions Deleted	0.44	3.01	Removed EFA
21	Total Triggered Events	0.36	4.66	Removed EFA
22	Total Decision Points	0.57	4.15	Removed EFA
23	Successfully Completed Science Quiz	3.34	0.83	
24	Delete Text From the Game	1.01	1.02	
25	End Editing Mode	1.42	1.41	
26	Successfully Completed Science Tasks	1.83	0.63	
27	End of Level Achieved	0.96	1.01	
28	Total Levels Edited	0.78	0.74	
29	End of Game Achieved	1.59	1.14	
30	Delete Session Notes	1.33	2.10	
31	Total Time In Game	0.56	2.18	Removed (a)
32	Map Code	0.55	3.12	Removed (b)
33	Total Data Stream	0.63	3.78	Removed (b)
34	Game Check Point Achieved	4.48	0.85	
35	Quiz Choices Added	0.72	0.98	
36	Quiz Saved	0.83	1.96	
37	Decision Point Triggered	1.31	0.96	
38	Total Decision Points Added	0.99	1.18	Removed EFA
39	Total Decision Points Deleted	0.66	0.85	Removed EFA
40	Decision Point Mapped	1.53	1.24	
41	Total End of Games Initiated	0.45	2.86	Removed EFA
42	Total Text Characters Added	0.45	2.71	Removed EFA
43	Total Text Characters Deleted	0.34	5.09	Removed EFA
44	Total Games Initiated	0.51	4.67	Removed EFA
45	Total Session Notes Made	0.60	1.01	Removed EFA

Table 5

Task Completion Probability

Task Number	Task Description	π_i	var (e_i)	var (τ_i)	ρ_{ii}
1	Science Based Task Editing	.38	.06	.01	.63
2	Initiated Game Play Mode	.41	.12	.03	.22
3	Initiated Editing Mode	.33	.08	.04	.37
4	Completed Science Content Based Task	.47	.10	.01	.49
5	Play Science Based Task Within Game	.44	.13	.04	.27
6	End Game Play Mode	.51	.16	.04	.21
7	Complete Science Task Development	.34	.08	.04	.36
8	Add Session Notes	.23	.02	.00	.03
9	Add Objects to Game	.15	.11	.01	.08
10	Delete Objects from Game	.27	.07	.00	.04
11	Add Text to the Game	.02	.02	.01	.23
12	Science Quiz Question Answered Correctly	.40	.08	.01	.57
13	Quiz Questions Added	.48	.17	.04	.20
14	Quiz Questions Deleted	.31	.09	.00	.01
15	Decision Point Added	.43	.10	.07	.38
16	Decision Point Deleted	.40	.12	.04	.26
17	Total Quiz Questions Added	.32	.18	.04	.11
18	Edit Decision Point	.35	.08	.06	.40
19	Quiz Notes Added	.24	.03	.00	.09
20	Total Quiz Questions Deleted	.11	.01	.00	.02
21	Total Triggered Events	.06	.06	.00	.00
22	Total Decision Points	.02	.02	.00	.23
23	Successfully Completed Science Quiz	.41	.06	.11	.68
24	Delete Text From the Game	.43	.12	.06	.29
25	End Editing Mode	.32	.08	.03	.34
26	Successfully Completed Science Tasks	.50	.01	.11	.53
27	End of Level Achieved	.44	.13	.06	.28
28	Total Levels Edited	.53	.17	.06	.24
29	End of Game Achieved	.17	.08	.07	.43
30	Delete Session Notes	.25	.03	.00	.18
31	Total Time In Game	.14	.11	.00	.07
32	Map Code	.07	.07	.00	.04
33	Total Data Stream	.02	.02	.00	.23
34	Game Check Point Achieved	.40	.08	.01	.57
35	Quiz Choices Added	.48	.17	.04	.20
36	Quiz Saved	.31	.09	.00	.01
37	Decision Point Triggered	.42	.11	.07	.38
38	Total Decision Points Added	.20	.12	.04	.26
39	Total Decision Points Deleted	.32	.18	.04	.11
40	Decision Point Mapped	.35	.08	.06	.40
41	Total End of Games Initiated	.12	.10	.00	.02
42	Total Text Characters Added	.13	.11	.00	.03
43	Total Text Characters Deleted	.05	.04	.00	.00
44	Total Games Initiated	.02	.02	.00	.21
45	Total Session Notes Made	.30	.18	.03	.17

Note. *P*i = 0.366, VAR (*e*) =4.212, VAR (τ) =51.051, P=0.95

The person item map displayed in Figure 5 provides subject scores and relative difficulty of items on a logit scale. The left side of the plot displays subjects' responses and the right hand shows the item difficulty. The item person map displays those subjects exhibiting the highest level of task completion at the top. Displayed at the bottom of the plot are respondents who least likely completed the tasks successfully. An approximate equivalent distribution, shown between the items and respondents, suggests tasks were of appropriate difficultly for the respondents. In addition, there seems to be adequate item coverage for all levels of ability as there not significant gaps within the person item map. Within the 2PLM item response model, Table 4 provides information regarding item parameters *a* and *b*, Fit statistics for the 2PLM model suggest adequate model fit for the data (χ^2=1.70, *df*=1, *p*=0.19).

Comparison of 2PLM model fit statistics and one-parameter logistic item response model (1PLM) fit statistics suggest that the 2PLM model was more appropriate. 1PLM model fit statistics resulted in a significant chi-square statistics suggesting a significant deviation from the expected results. The study did not consider a three-parameter logistic item response mode (3PLM) as the items are representative of tasks and guessing was not a feasible option. Thus, the study used a 2PLM model to develop population parameters and select tasks for later development into the artificial neural network and the Q-matrix under the IRT TRUE model in Table 5 (Dimitrov, 2009). Appendix B depicts the item characteristics curves and over all test information function.

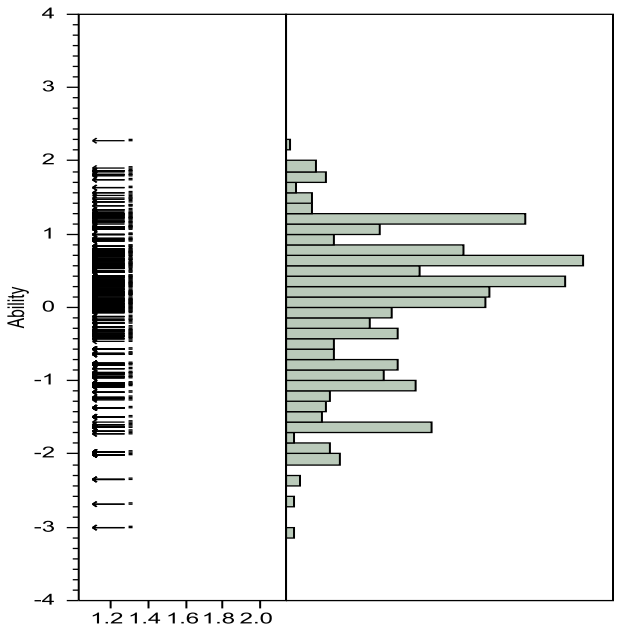


Figure 5. Item Person Map

Reliability

Estimation of reliability for the measured constructs used the Latent Trait Reliability Method (LTRM) (Dimitrov, 2012; Raykov, 2009; Raykov, Dimitrov & Asparouhov, 2010). This method (LTRM) provides superior estimation of internal reliability as it does not rely upon the assumptions associated with more common reliability methods such as Cronbach's alpha. Specifically, Cronbach's alpha requires essential tau equivalence and no correlated errors. Within the framework for latent variable modeling, score reliability developed as the ratio of the true-score variance to the observed variance (Dimitrov, 2012). Mplus code for LTRM is included in Appendix A. The reliability of the measured constructs is estimated at REL = 0.78, CI 5% [0.76-0.80],

SEM 2.32, CI 5% [2.27- 2.37]. The computed level of reliability is adequate for this type of measure. The study reported reliability based upon the current sample and not computed at the population level.

Task Attribute Matrix

The results of the Neural Network Analysis approach answer research question 3; *what theoretical mathematical / statistical model develops using an Artificial Neural Network to describe the interaction of the items, factors, and attributes as subjects design Serious Educational Games?* Results within this section suggest the rejection of the null hypothesis (H_0 R^2=0). Constructs such as this the relationship between attributes and tasks, in an exploratory study such as this reflects measure performance on tasks (Cronbach & Meehl, 1955; Embretson, 1983). Construct validity is the degree to which a scale measures the proposed trait is thought to measure. When such a test measures a trait, which is difficult to define such as in a cognitive diagnostic measure, multiple expert reviewers may rate individual pairings of attributes with tasks. Table 6 shows the independent relevance rating for the item task construct contained in this study.

Table 6

Relevance Rating for Each Task Attribute Pairing

		Reviewer 2	
		Weakly Relevant	Strongly Relevant
Reviewer 1	Weakly Relevant	18	10,29,34,35,36
	Strongly Relevant	15,16,24,25,28,30	1,2,3,4,5,6,7,8,12,13, 14,19,23,26,27,37,40

Note. Numbers correspond to individual tasks within the design of SEGs. Letters corrections to relevance grouping; A: WR x WR, B: WR x SR, C: SR x WR and D: SR x SR

Analysis of reviewer agreement of relevance suggests a task-attribute validity coefficient of 0.59. This level of task-attribute validity is adequate for an exploratory study such as this one. Equation 5 is a calculation of task-attribute validity using the coefficient of agreement *d*.

$$d = I_D / \sum_{I=1}^{n} I_{A\text{-}D} \quad (5)$$

A discrete latent attribute model was used to develop an understanding of the place of each cognitive attribute within the current model. This model allows for the modeling of cognitive weighting -via artificial neural network propagation weights- and for inferences about the hierarchical position of the cognitive attributes of the subjects. Within the models, the latent variables conceptualize as a vector of *0s* and *1s* for each subject. Zero indicates the absence of the trait and *1* indicates the presence of the trait. Table 7 illustrates the hypothetical attributes needed to complete the corresponding tasks.

More specifically to describe the model one can draw upon a similar model developed by Tatsuka (1995), where N examines and J binary task performances variables combine. A fixed set of K cognitive attributes are involved in performing the tasks. Thus, one can understand model parameters in the terms below,

X_{ij} = 1 or 0, indicating whether examinee *i* performed task *j* correctly;

Q_{jk} = 1 or 0 indicating whether attribute *k* is relevant to task *j*; and

α_{ik} = 1 or 0, indicating whether examinee *i* possesses attribute *k*.

Analysis of the Q-matrix assists with the standardization of the outputs by fixing term Q_{jk} to 1 prior to insertion into the matrix. The underlying reasoning for fixing term Q_{jk} equal to 1 is similar to the logic associated with the Linear Logistic Test Model (LLTM). From this development, it is important to understand the objective is to infer about the latent cognitive attributes developed via the artificial neural network model weightings. This is not to suggest traits the examinees do or do not possess but to aggregate the attributes along with suitable tasks to measure them. Note that the matrices are developed out of statistical estimations associated with task response parameters under a 2PLM. Table 7 also displays the odds of successful completion of the task. Table 8 illustrates the binary Q-matrix.

Table 7

Task Attribute Matrix (Q-Matrix)

Task Number	Task Name	Proposition of correct response on item *i*	Suggested Cognitive Attributes		
	Factor 1				
1	Science Based Task Editing	.37	Parity Judgment	Critical Reasoning	Retrieval
4	Completed Science Content Based Task	.47	Visual Attention	Critical Reasoning	Inference
5	Play Science Based Task Within Game	.44	Peripheral Localization	Attention Switching	General Cognition
7	Complete Science Task Development	.34	Reading	Inference	Visual Motor Coordination
12	Science Quiz Question Answered	.40	Reading	Inference	Critical Reasoning
23	Successfully Completed Science Quiz	.41	Reading	Inference	Critical Reasoning
26	Successfully Completed Science Tasks	.49	Reading	Inference	Visual Motor Coordination
27	End of Level Achieved	.44	Magnitude Quantification	Reading	Variable Interaction
29	End of Game Achieved	.36	Spatial Ability	Inference	Reading
34	Game Check Point Achieved	.40	Reading	General Cognition	Visual Attention
	Factor 2				
2	Game Play Mode	.41	Visual Attention	Retrieval	Peripheral Localization
3	Initiated Editing Mode	.33	Parity Judgment	Critical Reasoning	Retrieval
6	End Game Play Mode	.51	Reading	Retrieval	Critical Reasoning
8	Add Session Notes	.23	Reading	Verbal Production	Variable Interaction
10	Delete Objects from Game	.27	Visual Motor Coordination	Inference	Reading
24	Delete Text From the Game	.43	Visual Motor Coordination	Inference	Reading
25	End Editing Mode	.32	Reading	Retrieval	Critical Reasoning
28	Level Editing	.53	Peripheral Localization	Retrieval	Variable Interaction
30	Delete Session Notes	.25	Visual Motor Coordination	Inference	Reading
	Factor 3				
13	Quiz Questions Added	.48	Reading	Retrieval	Inference
14	Quiz Questions Deleted	.31	Reading	Retrieval	Critical Reasoning
19	Quiz Notes Added	.24	Reading	Retrieval	Verbal Production
35	Quiz Choices Added	.48	Inference	Estimation	Calculation
36	Quiz Saved	.31	Visual Attention	Reading	Estimation
	Factor 4				
15	Decision Point Added	.42	Variable Interaction	Estimation	Magnitude Quantification
16	Decision Point Deleted	.40	Estimation	Variable Interaction	Critical Reasoning
18	Edit Decision Point	.35	Reading	Inference	Visual Attention
37	Decision Point Triggered	.42	Peripheral Localization	Visual Attention	Attention Switching
40	Decision Point Mapped	.35	Visual Motor Coordination	Retrieval	Spatial Ability

Note. The positioning of the cognitive attributes is not meant to convey a hierarchical relationship between the attributes.

Table 8 illustrates the dichotomous coding for each of the cognitive attributes as related to tasks. Table 8 provides a listing of each cognitive attribute by number. A “1” indicates the presence of an attribute and a “0” indicates the absence of an attribute. Each task is limited to three cognitive attributes due to computational concerns in particular the lack of computing power within the context of this study. This seems to indicate that the lack of attributes is not a function of the methodology but a physical limitation imposed by a ceiling on computing power. Appendix C displays the cognitive attribute number and name.

Table 8

Q-Matrix

Number	Task / Attribute	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	Science Based Task Editing	0	0	1	0	0	0	0	1	0	0	1	0	0	0	0	0
2	Initiated Game Play Mode	0	0	0	0	0	0	0	0	1	0	1	0	0	0	1	0
3	Initiated Editing Mode	0	0	1	0	0	0	0	1	0	0	1	0	0	0	0	0
4	Completed Science Content Based Task	0	0	1	0	0	1	0	0	0	0	0	0	0	0	1	0
5	Play Science Based Task Within Game	1	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0
6	End Game Play Mode	0	0	1	0	0	0	0	0	0	1	1	0	0	0	0	0
7	Complete Science Task Development	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	1
8	Add Session Notes	0	0	0	0	0	0	0	0	0	1	0	0	1	1	0	0
10	Delete Objects from Game	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	1
12	Science Quiz Completed	0	0	1	0	0	1	0	0	0	1	0	0	0	0	0	0
13	Quiz Questions Added	0	0	0	0	0	1	0	0	0	1	1	0	0	0	0	0
14	Quiz Questions Deleted	0	0	0	0	0	0	0	0	0	1	1	0	0	1	0	0
15	Decision Point Added	0	0	0	1	0	0	1	0	0	0	0	0	1	0	0	0
16	Decision Point Deleted	0	0	1	1	0	0	0	0	0	0	0	0	1	0	0	0
18	Edit Decision Point	0	0	0	0	0	1	0	0	0	1	0	0	0	0	1	0
19	Quiz Notes Added	0	0	0	0	0	0	0	0	0	1	1	0	0	1	0	0
23	Successfully Completed Science Quiz	0	0	1	0	0	1	0	0	0	1	0	0	0	0	0	0
24	Delete Text From the Game	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	1
25	End Editing Mode	0	0	1	0	0		0	0	0	1	1	0	0	0	0	0
26	Successfully Completed Science Tasks	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	1
27	End of Level Achieved	0	0	0	0	0	0	1	0	0	1	0	0	1	0	0	0
28	Total Levels Edited	0	0	0	0	0	0	0	0	1	0	1	0	1	0	0	0
29	End of Game Achieved	0	0	0	0	0	1	0	0	0	1	0	1	0	0	0	0
30	Delete Session Notes	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	1
34	Game Check Point Achieved	0	0	0	0	1	0	0	0	0	1	0	0	0	0	1	0
35	Quiz Choices Added	0	1	0	1	0	1	0	0	0	0	0	0	0	0	0	0
36	Quiz Saved	0	0	0	1	0	0	0	0	0	1	0	0	0	0	1	0
37	Decision Point Triggered	1	0	0	0	0	0	0	0	1	0	0	0	0	0	1	0
40	Decision Point Mapped	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	1

Artificial Neural Network

The artificial neural network developed to describe the interconnection between the cognitive attributes and successful task completion arises from a series of interconnected nodes (neurons). The neurons develop the three district layers of the ANN- input, hidden, and output. The input layer provided no computational function but distribute stimulus into the neural network. For the purposes of this model, the tasks act as the input. The hidden layer represented by the cognitive attributes assigned to the tasks and the output layer consists of the success and failure probabilities. Figure 5 provides a generalized picture of the neural network used in this study.

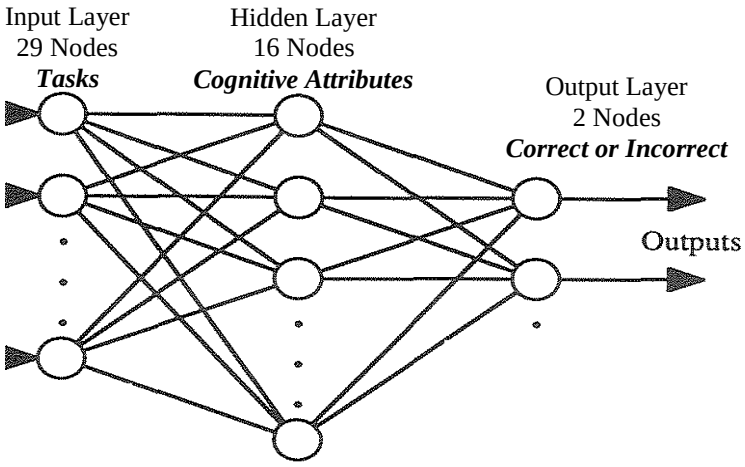


Figure 6. Overview of the Artificial Neural Network Topology.

The ANN used within the portion of the study was designed by SAS as part of the JMP 10 statistical discovery package. Training of the artificial neural network used a random 1/2 n split data approach similar to the validation method for the factor structure. Link weights initially consist of randomly weighted values. The weights are limited to random values within the range of $-2/\Omega$, $2/\Omega$ for neurons with Ω inputs (Gallant, 1993; Dawson & Wilby, 1997).

Post initialization of the network using the random weighting approach, the network was then trained by providing it (ANN) a number of examples from the 1/2 N data set (1/2 N = 77,120) illustrating how the ANN is to behave. Review of the results of the trained ANN with the calibrations set suggests an accurate behavioral predictor of subject success outcomes based on the cognitive attributes supplied. Table 10 and Table 11 provide key statistics regarding model fit. The training set shows a .86 and .78 r^2 for the prediction of correctly completing the tasks and incorrectly completing the task. These r^2 values suggest that the ANN model accounts for 86% and 78% of the variance around the sigmoid function used to develop the outputs. The generalized r^2 proves for the aggregation of the predictive ability of the network across the multiple outputs of *correct* and *incorrect*.

Review of Table 9 and Table 10, the ANN output for the test set of data suggests that the ANN model used for the test set is less able to predict the output states, *correct* or *incorrect* task completion ($\Delta\ r^2$=-0.13). Despite some loss in predictive power associated with the model, there is not a statistically significant difference in the r^2 values (t (2) =1.59, p=0.252, α=0.05). Given the lack of significance for chi-square, the model

adequately predicts subject outcomes using cognitive diagnostic approaches. When tested using the second set of data 1/2n the model is able to account for 77% of the variance for *correct* outcomes and 69% of the *incorrect* outcomes. Examination of the ΔRMSE (+0.02) term, there is a slight increase in the error term however this is not considered significant (t (2) =1.34, p=0.31, α=0.05). Review of the correlation coefficient r=0.85 suggests there is a strong linear relationship between the models.

Table 9

Neural Network Output (Training Set, 0.5 Holdback Validations

Neural Network	Correct	Incorrect
R-square	0.86	0.78
RMSE	0.19	0.21
Mean Abs Error	0.10	0.07
Generalized R-Square	0.82	

Table 10

Neural Network Output (Test Set, 0.5 Holdback Validations)

Neural Network	Correct	Incorrect	Average Change from Training
R-square	0.77	0.69	-0.02
RMSE	0.21	0.22	+0.015
Mean Abs Error	0.12	0.17	+0.06
Generalized R-Square	0.73		
Correlation Coefficient:	0.85		

An Artificial neural network derives propagation weights from random assignment to test set data (1/2n) for each of the proposed attributes. The weights represent the strength of signal propagation as the signal moves from node to node within the network. For clarity, the study has standardized ANN weights to 1.00, each subsequent weighting value developed from the standardized value. The attribute with the greatest weighting (largest likelihood to propagate) is perceptual binding (A13). Attribute 2 (A2) mental calculation, is the least likely to propagate. Table 11 displays the cognitive attribute name to the artificial neural network weighting.

Table 11

Neural Network Propagation Weights

Cognitive Attribute Number	Cognitive Attribute Name	Artificial Neural Network Weightings
1	Attention Switching	0.46
2	Mental Calculation (Arithmetic)	0.01
3	Critical Reasoning	0.30
4	Estimation (Numeral Sense)	1.00
5	General Cognition	0.01
6	Inference	0.20
7	Quantification	0.11
8	Parity Judgment	0.07
9	Peripheral Stimulus Localization	0.05
10	Reading	0.27
11	Retrieval	0.01
12	Spatial Ability	0.04
13	Variable Interaction (Perceptual Binding)	0.68
14	Verbal Production (Verbal Fluency)	0.14
15	Visual Attention	0.41
16	Visual Motor Coordination (Motor Control)	0.16

By combining neural network propagation weighting with the 2PLM probability of item completion, one can merge the two models and create a means to measure the

contribution each attribute makes to the overall task completion. Equation 6 represents the 2PLM.

$$P_i(\Theta) = \exp[Da_i(\Theta - b_j)] / 1 + \exp Da_i(\Theta - b_j) \quad (6)$$

By combining neural network propagation weighting Equation 7 with the 2PLM probability of item completion, one can merge the two models and create a means to calculate the propagation. Equation 7 represents the propagation weightings across the neural network.

$$X^j{}_n = \varphi\left(\sum_{l=0}^{Cn-1} w\hat{}il(n)(x\hat{}l(n-1)\right) \quad (7)$$

Equation 8 represents the combination of the two equations to represents the calculation of the probabilities of task completion related to each cognitive attribute.

$$\exp[Da_i(\Theta - b_j)] * \varphi\left(\sum_{l=0}^{Cn-1} w\hat{}il(n)(x\hat{}l(n-1)\right) / [1 + \exp[Da_i(\Theta - b_j)]] \quad (8)$$

Where *D* is the scaling factor equal to 1.70, (this approximates a normal ogive curve), a_i is the item discrimination, Θ is the subjects ability for success on the particular item. Through manipulation of these variables, one can calculate $P_{i,}$ the probability of correctly completing a task. When combined with the neural network model, φ is the non-linear activation function for the artificial neural network, $w^{il}{}_n$ represents the gradient decent along the training function, and $x^{l}{}_{n-1}$ represents the input to the hidden layers via

the cognitive attributes. Substitution for the expressions within the equation results in the following summary Equation 9 representing the probability contribution each cognitive attribute makes to the overall probability of task completion.

$$P_{Ai} = Pi\ (\Theta) * \varphi\ (y^{1}_{n}) \qquad (9)$$

Table 11 illustrates a modified Q-matrix with the addition of the probabilities for each of the attributes.

Table 11

Q-Matrix with Calculated Probabilities

Task Description	Pi	Attribute 1	Probability	Attribute 2	Probability	Attribute 3	Probability	Residual
Science Based Task Editing	0.38	Parity Judgment	0.03	Critical Reasoning	0.11	Retrieval	0.01	0.13
Completed Science Content Based Task	0.47	Visual Attention	0.19	Critical Reasoning	0.14	Inference	0.10	0.42
Play Science Based Task Within Game	0.44	Peripheral Localization	0.02	Attention Switching	0.21	General Cognition	0.01	0.23
Complete Science Task Development	0.34	Reading	0.09	Inference	0.07	Visual Motor Coordination	0.05	0.21
Science Quiz Question Answered	0.40	Reading	0.11	Inference	0.08	Critical Reasoning	0.12	0.31
Successfully Completed Science Quiz	0.41	Reading	0.11	Inference	0.08	Critical Reasoning	0.12	0.32
Successfully Completed Science Tasks	0.49	Reading	0.13	Inference	0.10	Visual Motor Coordination	0.08	0.31
End of Level Achieved	0.44	Magnitude Quantification	0.05	Reading	0.12	Variable Interaction	0.30	0.46
End of Game Achieved	0.37	Spatial Ability	0.01	Inference	0.07	Reading	0.10	0.18
Game Check Point Achieved	0.40	Reading	0.11	General Cognition	0.01	Visual Attention	0.17	0.28
Game Play Mode	0.41	Visual Attention	0.17	Retrieval	0.01	Peripheral Localization	0.02	0.19
Initiated Editing Mode	0.33	Parity Judgment	0.02	Critical Reasoning	0.10	Retrieval	0.01	0.12
End Game Play Mode	0.51	Reading	0.13	Retrieval	0.01	Critical Reasoning	0.15	0.29
Add Session Notes	0.23	Reading	0.06	Verbal Production	0.03	Variable Interaction	0.15	0.25
Delete Objects from Game	0.27	Visual Motor Coordination	0.04	Inference	0.05	Reading	0.07	0.17
Delete Text From the Game	0.43	Visual Motor Coordination	0.07	Inference	0.09	Reading	0.12	0.27
End Editing Mode	0.32	Reading	0.09	Retrieval	0.01	Critical Reasoning	0.10	0.18
Level Editing	0.53	Peripheral Localization	0.03	Retrieval	0.01	Variable Interaction	0.36	0.38
Delete Session Notes	0.25	Visual Motor Coordination	0.04	Inference	0.05	Reading	0.07	0.15
Quiz Questions Added	0.48	Reading	0.13	Retrieval	0.01	Inference	0.10	0.23
Quiz Questions Deleted	0.31	Reading	0.08	Retrieval	0.01	Critical Reasoning	0.09	0.17
Quiz Notes Added	0.24	Reading	0.07	Retrieval	0.01	Verbal Production	0.03	0.10
Quiz Choices Added	0.48	Inference	0.10	Estimation	0.48	Calculation	0.01	0.59
Quiz Saved	0.31	Visual Attention	0.13	Reading	0.08	Estimation	0.31	0.51
Decision Point Added	0.43	Variable Interaction	0.29	Estimation	0.43	Magnitude Quantification	0.05	0.51
Decision Point Deleted	0.40	Estimation	0.40	Variable Interaction	0.27	Critical Reasoning	0.12	0.80
Edit Decision Point	0.35	Reading	0.09	Inference	0.07	Visual Attention	0.14	0.31
Decision Point Triggered	0.42	Peripheral Localization	0.02	Visual Attention	0.17	Attention Switching	0.20	0.39
Decision Point Mapped	0.35	Visual Motor Coordination	0.06	Retrieval	0.01	Spatial Ability	0.01	0.07

CHAPTER FIVE

Discussion

The primary purpose of this dissertation was to design, validate, and establish a new methodological approach for the development of a cognitive diagnostic approach using large, derived data sets. The remainder of this chapter outlines the treatment and implications organized by research question.

Research Question 1

The results above clearly illustrate that there is an underlying clustering to the tasks contained in the design of SEGs. The results suggest the rejection of the null hypothesis, $\Lambda_i=0$ is appropriate as the factor loading is significant and above 0.30 for each item assigned to a factor. Research question 1; *what are the undying factors exhibited though the measurement of task items associated with subject development of science based Serious Educational Games?* Emergent factors developed from the exploratory factor analysis provide a starting point for the dimensional reduction and organization of the analysis of the task items. EFA reveals four orthogonal factors. Factors related to task items are Factor 1, *Science Processing*, Factor 2, *Game Control Actions*, Factor 3, *Evaluation and Assessment Development* and Factor 4, *Games Logic*. Two of the four factors confirm previous results seen in the pilot study, specifically, Factor 1 and Factor 2. This is an expected result, as these two factors are consistent

across SEG based upon science content. Form a logical point of view, game control functions are the most fundamental interaction possible within any video game and, would certainly be present in an SEG. Along these lines, the science processing would be the most fundamental interaction possible with the science content of a game. Meaning, that a subject designing a science based game would be required to process science tasks or science concepts in order to drive the game forward. The remaining two factors *Evaluation and Assessment* and Game *Logic*, where not discovered within the pilot study. However, the factors seem specific to the game design process. CFA also provides a means to establish that the four factors are locally independent and, subsequently, it is possible to use IRT analysis to parameterize and model components. PCA provides a secondary confirmation of the hypothetical number of factors as analysis of eigenvalues reveals four factors using the root >1 criteria. Rotated solutions also reveal a simple structure with four linearly independent factors. Cross-validation of the EFA using 1/2n CFA suggests that the suggested data structure and factor loading is indicative of four latent traits or factors. Examples of task items loading on each factor is, Factor 1, *Science Processing,* to complete science task development and science quiz questions correctly answered. Initiated editing, and delete objects from the game, exemplify Factors 2, *Game Control Actions*. Quiz choices added, and quiz questions deleted exemplify Factor 3, *Evaluation, and Assessment Development.* Examples for the final factor, Factor 4, *Game Logic*, are Decision Point Triggered and Decision Point Mapped. It is important to understand that the factor analysis primarily serves as a means to organize the externalized actions of the subjects as they design SEGs. These factors serve as an

organizational structure for the underlying cognitive attributes as these attributes cross multiple factors. However, the factors do aid in the conceptualization of relationships and organization of the input nodes within the ANN.

The factor analysis also provided a means to remove tasks that would degrade the data structure of the analysis. Poor factor loading resulted in the removal of eleven factors. These factors were primarily on tasks related to aggregation of task items across multiple tasks. Examples of some of these items are, total quiz questions added and total decision points added. The aggregate nature of the task items confounds the factor analysis due to their correlation across multiple items that are not necessarily related. As part of the confirmatory analysis, the study used the latent trait reliability method as a means to establish task internal reliability. Overall reliability results suggest that the task acts as an internally consistent measure of task competition related to each of the four factors. Thus, this internal consistency links the clustering of the factors to the tasks. The factors provide a logical means to develop task relationships as a function of complex task completion. Due to the aggregation of the simple tasks, analysis of the larger complex task is possible via the ANN.

Research Question 2

The 2PLM IRT model was tested for data fit using the computer program JMP 10.0 (SAS, 2012). As a fit statistic, JMP reports a standardized residual that approximates a Gaussian distribution. Values that exceed 2.0 under parameter (b) indicated misfit at an alpha of .05. Within this study, item fit statistics range from 0.36 to 4.46 prior to the removal of items. Post removal of items, the range adjusts from 0.36 to 2.24. This

indicates that the 2PLM fits the data. The fit of the data to the 2PLM is of importance for development of the IRT-True score and parameterization of population level statistics. Estimated expected item scores, π_j, develop through an approximate evaluation of the empirical estimate of item difficulty (b) and discrimination (a). True-score measures provide substantive information and allow for identification of the most difficult task and easiest tasks. The most difficult task of the study is the addition of game notes. One would expect this task (add session notes) to be the most difficult as the number of integrated attributes for this particular task would be greater given the complexity of the task. Add session notes, is the least difficult task. From a substantive point of view, this is keeping with the overall perceived difficulty of the tasks. Overall estimation of reliability of task measures at the population level is 0.95, which is considerably higher than the individual internal reliability calculated at the sample level (0.78). The difference in reliably estimates may be due to natural variance in the sample level statistic.

The analysis of task items using IRT provided a means to select and order tasks within a statistical model. Educators and psychologists accomplished assignment of the cognitive attributes to each parameterized task via review. Category D represents task attribute alignments reviewers rated as strongly relevant. Items not placed into category D by the each of the reviewers showed mixed or weak relevance, i.e. one reviewer rated the task-attribute alignment as strongly relevant and one reviewer rated the same task-attribute alignment as weakly relevant or both reviewers rated each task attribute alignment as weakly relevant. This is not to suggest that the task-attribute alignments were not relevant but rather that the task attributes alignment was not as strong as other

combinations. However, given the unaccounted residual probabilities (error) within the quantified Q-matrix, lack of alignment is an expected outcome. This outcome "residual" is due to missing cognitive attribute accounting within the current model. One of the expert reviewers obtained their Ph.D. in Educational Psychology and currently works for the Social Science Research Institute at Duke University, working on projects related to STEM education of at risk youth. The second expert reviewer obtained their Ph. D. in school psychology from the University of Illinois-Urbana-Champaign and currently works within a large urban school district in the Mid-Atlantic region of the United States.

Attribute assignment is critical to the process. Given the exploratory nature of the study, assignment of attributes is tentative (hypothetical) and contingent upon literature, expert review, and emergent patterns via data mining. It is also important to note that this process is similar but not exactly, the same as cognitive diagnostics in the traditional psychology approach. The differences arise out of the way, in which the task-attribute combination is developed. Typically, within the development of cognitive diagnostic approaches the task-attribute relationship is far less complex and much more isolated than within the educational environment. However, given the need for more complete information regarding subject learning, a cognitive diagnostic approach embedded in an analytics approach. While this analysis occurs more quickly and with more tasks and fewer attributes, the information garnered from each individual attribute is less. However, overall attribute mastery patterns, including missing attributes that become visible, are thus viable targets for intervention and instruction. Monotonic behavior of the attributes

within the neural network suggests from a substance point of view the attributes are correctly assigned.

Early on within the development of the cognitive diagnostic models, construct-oriented designs suggest use in the assignment of attributes to tasks. Analysis of the literature suggests that there are several suggested attributes, which underlie tasks items. The outcomes of this process are the development of the Q-matrix. Several analyses quantify and parameterize the Q-matrix. Research question 2; *what are the cognitive attributes that underlie the design of Serious Educational Games?* Initial analysis centered on the development of task outcomes using dichotomous nominal quantifiers with *0* representing failure to complete the task and *1* representing successful completion of the task. Parameterization of the tasks post factor analysis occurred using a 2PLM. Review of parameter *a* (discrimination) and parameter *b* (difficulty) results suggests that several items should be removed for poor metrics. The items removed from the analysis are not contingent upon any specific attribute or task within the game design process. The removal of items aided in the development of ANN model fit and increased the predictive nature of the network. The weighting parameters also provide a means to establish individual attribute probabilities and the individual contribution of the attribute to the overall probability of task completion. This development of increased resolution via the addition of individual attribute probabilities use standardized ANN weightings. The strength of this least squared approach allows for examination of both general and local attribute maxima and minima. As with the pilot study, the use of the neural network allows for the analysis of the item attribute relationships with fewer items.

Research Question 3

The computational cognitive model (ANN) obtained information related to the theoretical and mathematical / statistical model of cognition related to science based Serious Educational Game design (RQ3). Hypothesis 3 (H3), is thus rejected. More specifically, research question 3 and the associated hypothesis validate the model. The ANN model exhibits good fit, and approximates human learning related to the design of the SEGs using science content. The processing and design tasks are divided up via factor analysis and input into the network using 29 input vectors on task as a time in random order. The network receives each task in a factor before any input sets repeat. Each experimental run began with a new set of weights. Each weight was bound to ensure sustentative outcomes. In each examination of the tasks and attributes, the network was trained on some combination of problem examples for a fixed number of iterations (I=1000). Model fit dictated the number of iterations resulting in convergence. Good model fit is suggestive of a computational-cognitive model describing the underlying cognitive attributes activated while designing SEGs ($H_{0,}$ R^2=0). Through the introduction of tasks as input nodes and attributes as the hidden nodes, it becomes possible to create a sophisticated model of cognition relating to science processing and the design process. Given that the development of science process and game design is very complex and poorly understood, an incremental approach such as this one produces high quality results. Attributes developed via the conjunctive cognitive diagnostic model provide a view of possible attributes. By imposing structure upon the order of attribute introduction during training, the network learner (simulated subject) focuses on the characters of a

smaller number of task types per attributes. As the number of attributes increases, the overall probability of task completion increases.

Initially the network trained on data sets designed to complete the tasks associated with each factor. Subsequent runs of the model focused on the use of novel data sets to provide a test of the ANN ability to complete tasks. Due to the essential unidimensional nature of the factors, there was no overlapping of tasks; however, individual attributes did overlap across tasks. The overlap of the cognitive attributes helps to explain non-linear outcomes associated with the learning. The model correctly completed tasks a significant portion of the time thus validating the model and creating four predicative models of subject learning using Bayesian statistical models. Figures 7 through 10 provide an overview of each of the models. Each of the models illustrates the relationship between the tasks and attributes. Model outputs can be manipulated increasing and decreasing attributes as a function of interventions with task probabilities of task success as the outcome. The output node labeled 1 in each figure represents the cumulative probability of the task grouping success given the full complement of the cognitive attributes. An interesting point to note is when the ratio of attributes (hidden nodes) to tasks is relatively high, the odds of success increase. This may be due to the distribution of the cognitive load between various attributes. The weightings and the contribution of the attribute to the overall probabilities of success evidence the distribution of the cognitive load. However, with manipulation of the cognitive attribute distribution, one can experiment with the role each attribute plays in science process and the design of SEGs. As

identification of future task-attribute combinations increase, the model predictive power also increases.

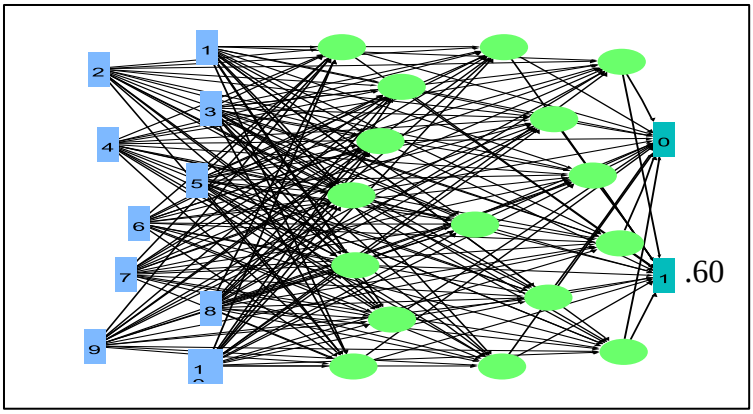


Figure 7. Factor 1 Neural Network Model.

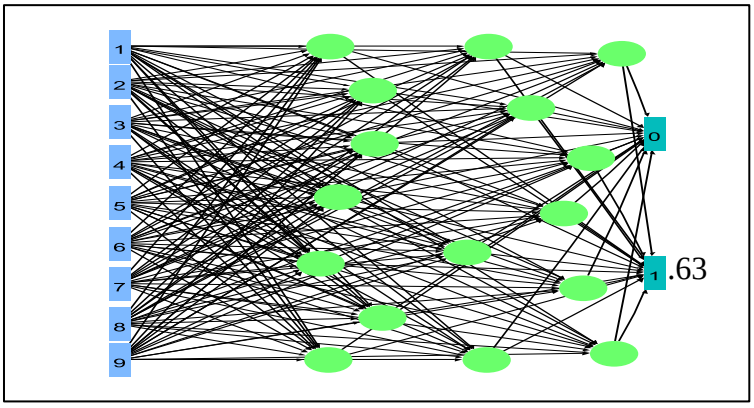


Figure 8. Factor 2 Neural Network Model.

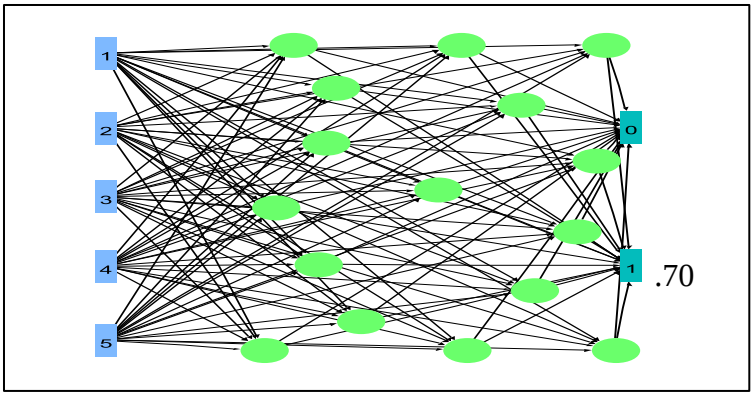


Figure 9. Factor 3 Neural Network.

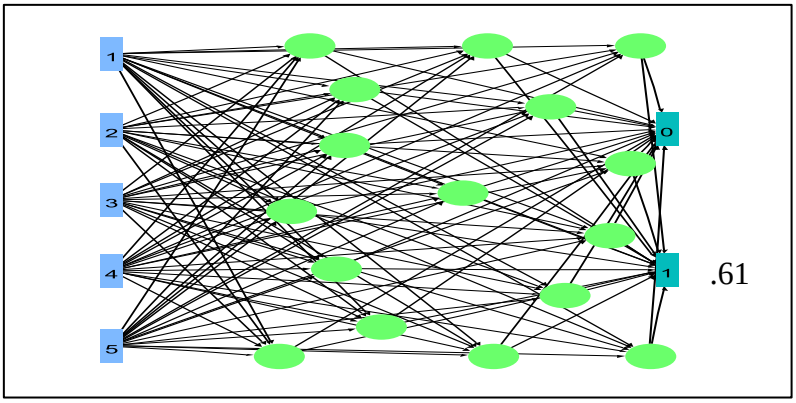


Figure 10. Factor 4 Neural Network.

Conclusion

Video games pay a substantial role in our culture as almost 70% of Americans have played some sort of video game and do so regularly (Kenny & McDaniel, 2011). Studies of video games suggest they dramatically enhance and alter a wide range of cognitive traits to enhance hand-eye coordination, reaction times, and mental rotation. Given the ability of the cognitive diagnostics and this novel measurement, technique to identify and establish parameters for task-attribute combination allows for enhanced targeting of attributes. While it is difficult to predict the future of technology, video game development can take many possible directions. Some of these changes will result from increases in graphic processing, perspective, realism, and speed. The increased level of detail and speed offers a smoother and more realistic experience and in turn, more complete transference to the real world. In turn, this realistic experience develops the basis for a Serious Educational Game assessment system by allowing realistic task presentation for subjects. This study has developed a justification for combining and developing two distinct areas of research related to subject learning. The first is the use of cognitive diagnostic approaches to assess subject learning as it relates to the cognitive attributes used during science processing. The second area is an examination and modeling of the relationship between attributes as propagated in an artificial neural network. The literature presented in this dissertation work integrates work from multiple filed areas. Fields represented in this work range from science education, educational psychology, measurement, and computational psychology.

Implications for science education. Science education will benefit from the expansion in the measurement capabilities in the field by providing a novel means to assess subject understanding and modeling of subject learning via cognitive attributes. This also sets the conditions for the enhanced understanding of higher-order cognition with fewer attributes. These models can lead to more specific targeting of subject learning increasing the effective use of SEGs as a means for teaching and assessing subjects. The proposed study also gives cognitive psychology a new way to evaluate conceptual learning outcomes.

Effective targeting of underlying cognitive attributes can result in reduction of the disjunction that exists between cognitive psychology and education and measurement. Analysis of the Attribute Mastery Patterns (AMP) and related factors yields far more information regarding potential areas of poor subject performance than traditional assessment analysis. The information garnered from this process can inform instructional approaches in the design process, which integrates cognitive psychology at the practitioner level of implementation. More specifically, it is possible to develop responsive SEGs using virtual environments. Measurement Environments using Responsive Cognitive Immersion (MERCI) would develop out of an IRT like computer adaptive testing model. Within a traditional IRT (simplified) testing model, subjects are presented with the items at various levels designed to measure a specific Θ. Based upon subject responses as a function of item difficultly questions are adjusted in an effort ascertain the subjects Θ. Within the MERCI model, Θ is a function of A_p or the probability estimate of someone with a particular cognitive attribute pattern successfully

solving a specific task as an identified problem set. This may provide a means to successfully measure - in real-time- subject outcomes not based on Θ but based upon individual attribute matrixes in A_p.

Development of this particular mode of assessment would assist science educators in the placement and targeting of particular types of problems to assist in subject cognitive development and development of novice STEM participant profiles to be more like expert profiles. As the profiles develop, one would expect an increase in the number and selection of STEM based courses in addition to increase scientific literacy. More importantly revised science curricula can develop to target key attributes within the hierarchy of cognition and help to develop all attributes downstream from the key attribute. While science educators have not implemented these changes in science curriculum, other researchers have employed a similar version of curriculum design through 'learning trajectories' in mathematics with success (Confrey, Maloney, Nguyen, Mojica, & Myers, 2009). For example, when discerning how to teach rational number reasoning, it was determined that the concept upon which the other concepts rested (in our case the most propagated attribute) was the understanding of dividing and sharing, shared by many children naturally. Potential examples for science reasoning stems from the design process itself: specifically, subjects must first learn how to isolate potential variables through experimental methods. This would help to ensure that subjects at all levels would benefit. In particular, subjects at the upper- and lower-levels of the learning outcomes continuum would benefit most, as specific targeting of attributes would provide the greatest probability of gains in scientific processing skills.

Limitations. Research on human cognitions has taken place within many disciplines and levels from the biological to the behavioral observation. Many of the original studies establishing the connection between cognitive attributes and traits were performed in isolation. The isolation of the attributes and tasks provides a means to control for irrelevant data and increases the generalizability of results. This study does not allow for the isolation of attributes but rather the functioning of attributes within the natural ecology of the classroom providing a holistic view of learning. This ecology produces confounding variables, which interfere with the assignment of attributes to tasks and requires careful consideration prior to assignment. Further studies are required to develop the relationships between attributes and tasks and the effect of the natural ecology of cognitive attribute functioning.

A second limitation of the study is the difficultly in differentiating subject play behaviors with design behaviors. Often the two (play and design) intertwine as subjects "test" their video game designs. In particular, the design process itself outlines this behavior (testing) as a key feature of the process. Play and development by subjects is a key consideration in as play and testing have a place in the design process. However, controlling for this behavior would result in a much less complicated data structure and a reduction in the dimensionality of the data. The connection between the two on the surface seems to be consequential and bears further study.

The third limitation is the difficultly of model validation. Meaning that the quality of the model is only, as good as the variables and known used to generate the model. As with any model, there are several points within the development of the ANN model

where trait estimation is necessary because of this estimation of error terms and variance become a key consideration along with the means to measure them. Further to this point, the task-attribute relationship coefficient provides a relatively uncomplicated way to assign relevance for an exploratory study such as this one, future studies should include the use of interclass-correlation for analysis of the reliability task-attribute relationships.

Future Work. Computational studies avoid many of the difficulties associated with traditional studies using live subjects. Computational models can bridge disciplinary boundaries and provide linkage to wider sets of knowledge and data. Combining computational models with larger computer generated data sets and data-mining techniques provide a means to examine transient trends more easily overlooked in conventional studies. Secondly, simulations and models can minimize the impacts of phenomena without controlling for it. Thirdly, simulation can appear to compress time making longitudinal investigation possible that otherwise would not be possible given material limitations.

Much like the arguments above as they relate to computational models, SEGs assessments can provide many of the same benefits when combined with computational models of subject learning. Whereas the benefits of video games to assess cognition and other skills are undeniable, the work within this area is in its infancy. In particular, it is necessary to identify the wide variety of attributes used, linking the attributes to the overlying facets, learning trajectories, and identify the data mining techniques, which allow researchers to analyze the vast data generated from these techniques. There are multiple directions for the future research. First, this paper focuses on establishing and

“diagnosing’ particular cognitive attributes as they relate to science processing tasks. There is a need for further research to investigate how these diagnostic scores inform instruction at the individual subject level. The development of a cognitive attribute assessment using real-time SEG based assessments is possible using a variety of conventional techniques. Analysis of items can result in the development of a Q-matrix, which can play an important role in the creation of subject cognitive profiles and leads to more efficient presentation of tasks within the Serious Educational Game environment. From these profiles, subjects with the assistance of science teachers and curriculum designers can engage in cognitive process by selectively directing attention to different aspects of the Serious Education Game environments. More importantly, evidence for the development of cognitive skill sets is necessary for players to conduct complex cognitive processes such as reading explicit information, inductive reasoning, and problem solving in a flexible process.

APPENDICES

APPENDIX A

Software Coding for Statistical Analysis

Mplus Code for Confirmatory Factor Analysis

```
TITLE: CONFIRMATORY FACTOR ANALYSIS
DATA:FILE IS "C:/COGDI.DAT";
VARIABLE: NAMES ARE Y1-Y45;
 USEVARIABLES ARE Y1-Y45;
 CATEGORICAL ARE Y1-Y45;
MODEL: Factor 1 BY Y1 Y4 Y5 Y7 Y12 Y23 Y26 Y27 Y29 Y34;
 Factor 2 BY Y2 Y3 Y6 Y8 Y9-Y11 Y24 Y25 Y28 Y30-33;
 Factor 3 BY Y13 Y14 Y19 Y35 Y36;
 Factor 4 BY Y15 Y16 Y18 Y37 Y40;
 F1 WITH F2 @0;
 F1 WITH F3 @0;
 F1 WITH F4 @0;
 F2 WITH F3 @0;
 F2 WITH F4 @0;
 F3 WITH F4 @0;
OUTPUT: STANDARDIZED MODINDICES;
```

Mplus Code for Latent Trait Reliability Estimation

```
TITLE: LTRM
DATA: FILE IS "C:/CODDI.DAT";
VARIABLE NAMES ARE X1-X45;
USEVARIABLES ARE X1-X19 X23-37 X40;
MODEL: ETA BY X1*(B1)
 X2-X19(B2-B19) X23-X37(B23-B19) X40(B40);
 ETA@1;
 X1-X19 X23-37 X40 (EV1-EV19 EV23-EV37 EV40);
MODEL CONSTRAINT:
NEW(REL SEM TVAR VAR XVAR);
TVAR = (B1+B2+B3+B4+B5+B6+B7+B8+B9+B10+B11+B12+B13+B14+B15+B16+
B17+B18+B19+B23+B24+B25+B26+B27+B28+B29+B30+B31+B32+B33+
B34+B35+B36+B37+B40)**2;
EVAR = EV1+EV2+EV3+EV4+EV5+EV6+EV7+EV8+EV9+EV10+EV11+EV12+
```

```
EV13+EV14+EV15+EV16+EV17+EV18+EV19+EV23+EV24+EV25+
EV26+EV27+EV28+EV29+EV30+EV31+EV32+EV33+EV34+EV35+
EV36+EV37+EV40;
REL = TVAR / (TVAR+EVAR);
XVAR = TVAR+EVAR;
SEM=SQRT(XVAR*(1-REL));
ANALYSIS: BOOTSTRAP = 5000;
OUTPUT: CINTERVAL(BCBOOTSTRAP);
```

Note. The Mplus code presented here is originally modified from Mplus code presented by Dimitrov (2012) during Educational Research 827 Development and Validation of Assessment Scales at George Mason University

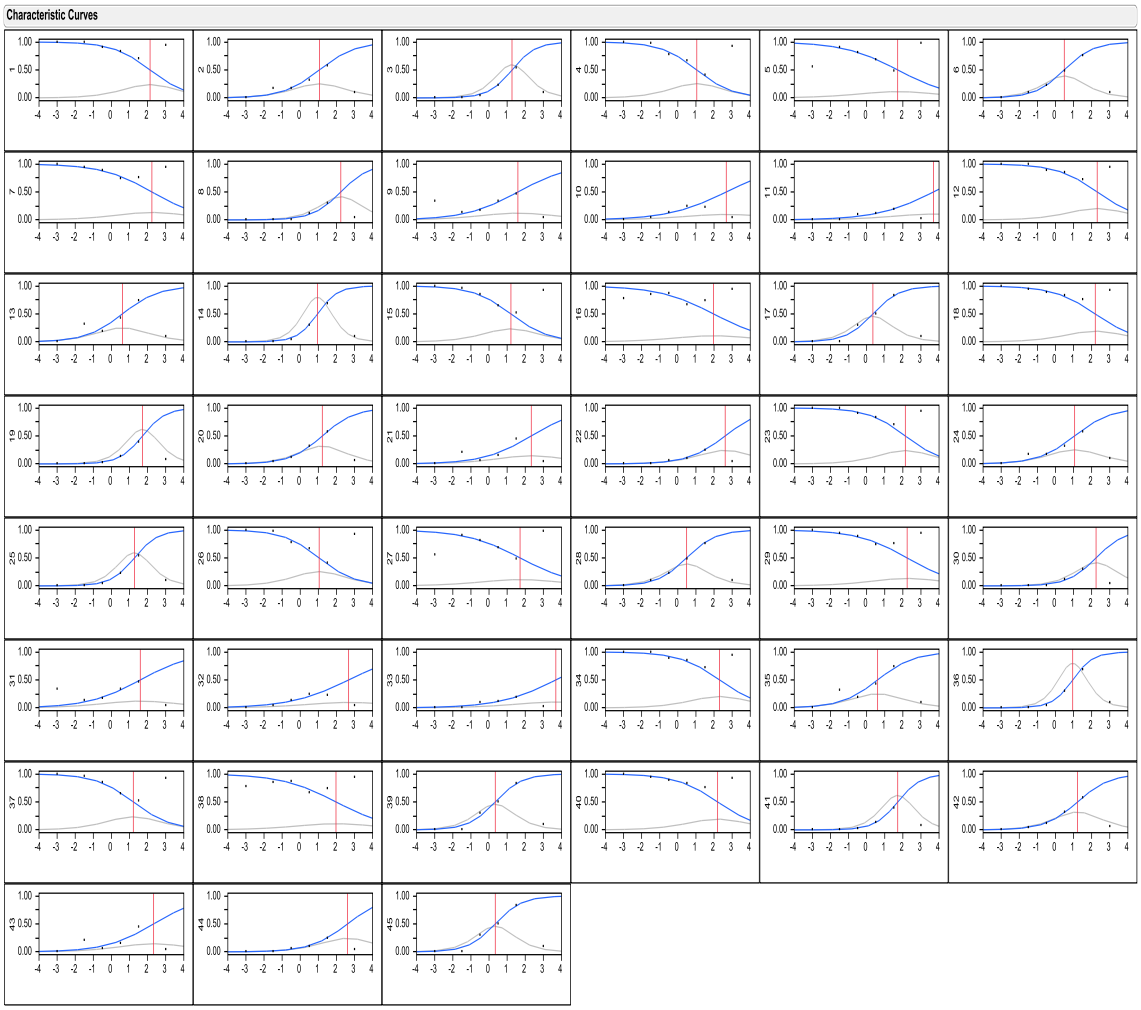
SAS Code for Neural Network Analysis

```
Neural(Y (: Name ("0"), Name ("1")),
X (: Name ("1"),
: Name ("2"),
: Name ("3"),
: Name ("4"),
: Name ("5"),
: Name ("6"),
: Name ("7"),
: Name ("8"),
: Name ("9"),
: Name ("10")),
Missing Value Coding (0),
Validation Method (Holdback, 0.5),
Fit (NTanH (16)));
```

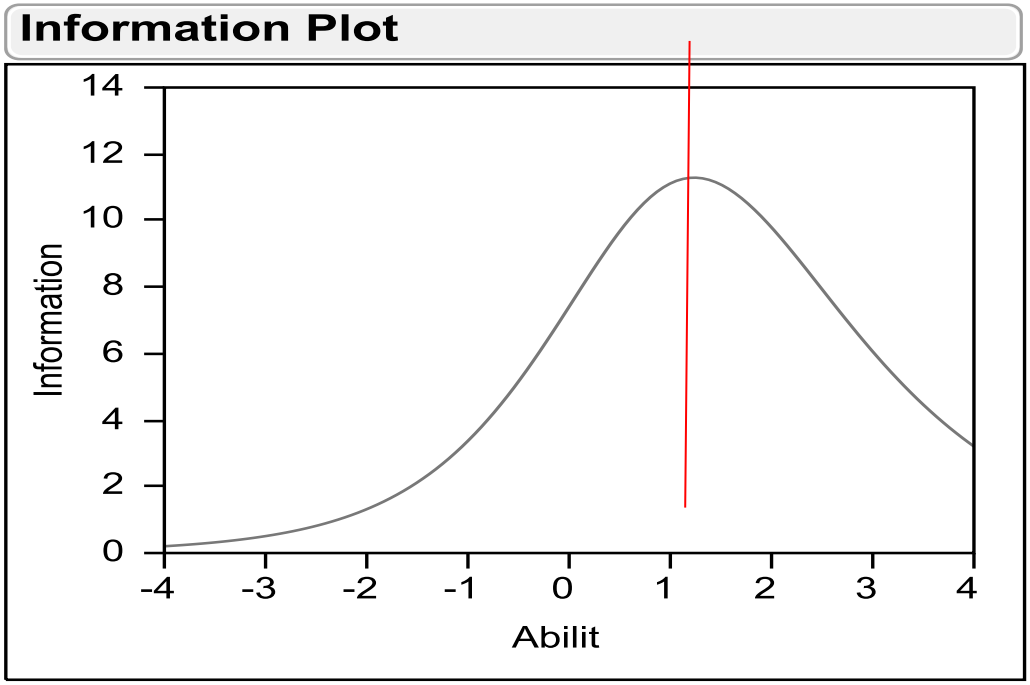
APPENDIX B

Item and Assessment Characteristic Curves

Item Characterize Curve Items 1 - 45



Test Infromation Function



APPENDIX C

Attribute Number and Name List

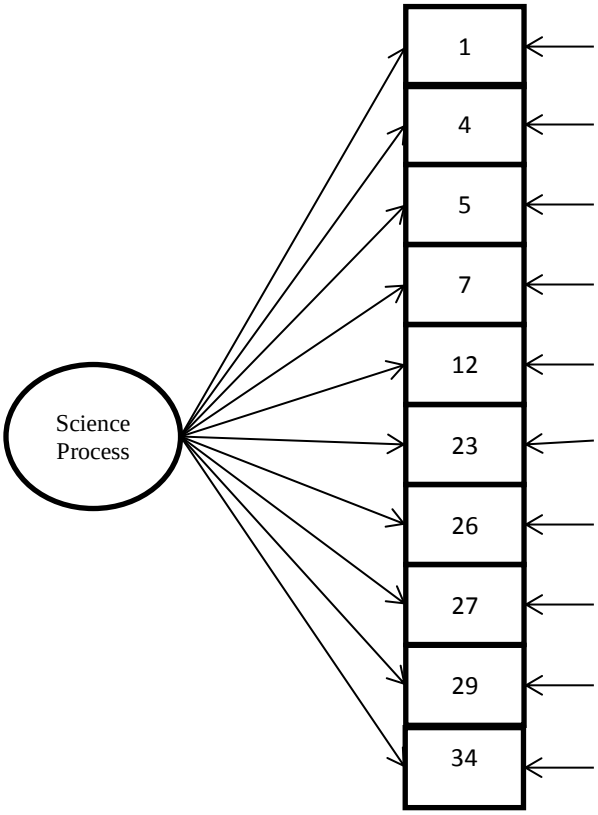
Attribute Listing

Number	Name
1	Attention Switching
2	Mental Calculation (Arithmetic)
3	Critical Reasoning
4	Estimation (Numeral Sense)
5	General Cognition
6	Inference
7	Quantification
8	Parity Judgment
9	Peripheral Stimulus Localization
10	Reading
11	Retrieval
12	Spatial Ability
13	Variable Interaction (Perceptual Binding)
14	Verbal Production (Verbal Fluency)
15	Visual Attention
16	Visual Motor Coordination (Motor Control)

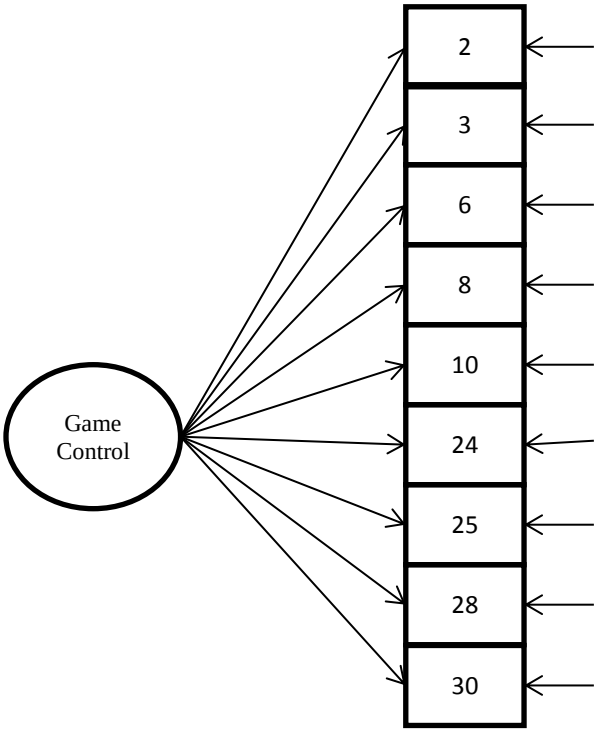
APPENDIX D

Confirmatory Factor Analysis Model

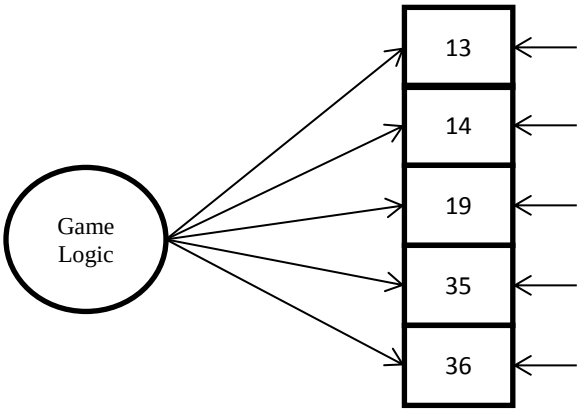
Factor 1 Model



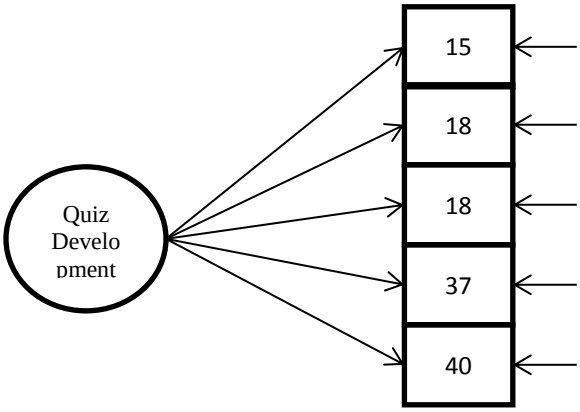
Factor 2 Model



Factor 3 Model



Factor 4 Model



APPENDIX E

IRB Approval

REFERENCES

Abell, S., & Lederman, N. (2007). *The handbook of research on science education*. Mahwah, N.J.: Lawrence Erlbaum Associates.

Alfieri, L., Brooks, P. J., Aldrich, N. J., & Tenenbaum, H. R. (2011). Does discovery-based instruction enhance learning? *Journal of Educational Psychology*, *103*(1), 1.

Allik, J., Tuulmets, T., & Vos, P. G. (1991). Size invariance in visual number discrimination. *Psychological Research*, *53*(4), 290-295.

Almond, R., DiBello, L., Moulder, B., & Zapata-Rivera, Z. (2007). Modeling diagnostic assessments with Bayesian networks. *Journal of Educational Measurement*, *44*(4), 341-359.

Annetta, L.A. (2010). The "I's" have it: a framework for Serious Educational Game design. *Review of General Psychology*, *14*(2), 105-112.

Annetta, L. A., Folta, E., & Klesath, M. (2010). *V-Learning: Distance education in the 21st century through 3D virtual learning environments*. New York, NY: Springer.

Annetta, L., Minogue, J., Holmes, S., & Cheng, M. (2009). Investigation the impact of video games on high school students' engagement and learning about genetics. *Computers & Education*, *53*(1), 74-85.

Annetta, L.A. (2008). *Serious educational games: From theory to practice*. Amsterdam, The Netherlands: Sense Publishers.

Annetta, L. (2008). Video games in education: why they should be used and how they are being used. *Theory into Practice, 47*(3), 229-239.

Arsalidou, M., & Taylor, M. J. (2011). Is 2+ 2= 4? Meta-analyses of brain areas needed for numbers and calculations. *NeuroImage, 54*, 2382-2393.

Baden, M. (2008). From cognitive capability to social reform? Shifting perceptions of learning in immersive virtual worlds. *Research in Learning Technology*, *16*(3), 151-161.

Bandura, A. (2006). *Guide for construction self-efficacy scales*. Greenwich, CT: Information Age Publishing.

Barbey, A. K., Koenigs, M., & Grafman, J. (in press). Dorsolateral prefrontal contributions to human working memory. *Cortex*.

Bateman, C., & Nacke, L. E. (2010, May). The neurobiology of play. In *Proceedings of the International Academic Conference on the Future of Game Design and Technology* (pp. 1-8). Vancouver, BC: ACM.

Beckmann, C. (2012). Modeling with independent components. *NeuroImage, 62*(2), 891-901.

Berglund, B., Rossi, G. B., & Wallard, A. (2012). *Measurement across physical and behavioral sciences*. New York, NY: Taylor and Francis Group.

Bernard, R. M., Abrami, P. C., Borokhovski, E., Wade, C. A., Tamim, R. M., Surkes, M. A., & Bethel, E. C. (2009). A meta-analysis of three types of interaction treatments in distance education. *Review of Educational Research, 79,* 1243-1289.

Betemps, E., & Baker, D. (2004). Evaluation of the Mississippi PTSE scale-Revised using Rasch Measurement. *Mental Health Services Research, 6*(2), 117-125.

Bhatt, M. (2012). Evaluation and associations: A neural-network model of advertising and consumer choice. *Journal of Economic Behavior and Organization*, *82*(1), 236-255.

Bingham, E., Kuusisto, J., & Lagus, K. (2002, August). ICA and SOM in text document analysis. In *Proceedings of the 25th annual international ACM SIGIR conference on Research and development in information retrieval* (pp. 361-362). Tampere, Finland: ACM.

Binney, R. J., Embleton, K. V., Jefferies, E., Parker, G. J., & Ralph, M. A. L. (2010). The ventral and infer lateral aspects of the anterior temporal lobe are crucial in semantic memory: evidence from a novel direct comparison of distortion-corrected fMRI, rTMS, and semantic dementia. *Cerebral Cortex, 20,* 2728-2738.

Birn, R. M., Kenworthy, L., Case, L., Caravella, R., Jones, T. B., Bandettini, P. A., & Martin, A. (2010). Neural systems supporting lexical search guided by letter and semantic category cues: a self-paced overt response fMRI study of verbal fluency. *NeuroImage, 49,* 1099-1107.

Bishop, M. (1995). *Neural networks for pattern recognition.* London, England: Oxford University Press.

Bond, C. E., Philo, C., & Shipton, Z. K. (2011). When there isn't a right answer: Interpretation and reasoning, key skills for twenty-first century geoscience. *International Journal of Science Education, 33,* 629-652.

Borgen, F. H., & Barnett, D. C. (1987). Applying cluster analysis in counseling psychology research. *Journal of Counseling Psychology, 34,* 456-468.

Breuer, J. & Bente, G. (2010). Why so serious? On the relation of serious games and learning. *Edamos Journal for Computer Game Culture, 4*(1), 7-24.

Briggs, D. C., & Alonzo, A. C. (2012). The psychometric modeling of ordered multiple-choice item responses for diagnostic assessment with a learning progression. *Learning Progressions in Science, 11*(1), 293-316.

Bruner, J. S., Goodnow, J. J., & Austin, G. A. (1986). *Study of thinking*. Brunswick, NJ: Transaction Publishers.

Buchweitz, A., Mason, R., Tomitch, L., & Just, M. (2009). Brain activation for reading and listening comprehension: An fMRI study of modality effect and individual differences in language comprehension, *Psychology & Neuroscience, 2*(2), 111-123.

Cabeza, R., & Moscovitch, M. (2013). Memory systems, processing modes, and components functional neuroimaging evidence. *Perspectives on Psychological Science, 8*(1), 49-55.

Caporale, N., & Dan, Y. (2008). Spike timing-dependent plasticity: a Hebbian learning rule. *Annual Review of Neuroscience, 31,* 25-46.

Carpenter, G. A. (1989). Neural network models for pattern recognition and associative memory. *Neural Networks, 2*(4), 243-257.

Cattell, R. (1966). The scree test for the number of factors. *Multivariate Behavioral Research, 1*(2), 245-276.

Chang, H. (2012). Local fisher discriminate analysis based manifold-regularized SVM model for financial distress predications. *Expert Systems with Application, 39,* 3855-3861.

Charsky, D. (2010). From edutainment to serious games: A change in the use of game characteristics. *Games and Culture, 5*(2), 177-198.

Chen, C., Wong, K., Leung, K., & Kwan, R. (2012). An enhanced e-assessment system for the acquisition of Putonghua. *Communications in Computer and Information Science, 3*(2), 45-58.

Choi, H. J. (2010). *A model that combines diagnostic classification assessment with mixture item response theory models* (Doctoral dissertation). Retrieved from http:// http://ugakr-maint.libs.uga.edu/handle/123456789/7436

Clegg, A. (1979). Craftsmen and the origin of science. *Science & Society, 43*(2), 186-201.

Clements, D., & Sarama, J. (2011). Early childhood teacher education: the case of geometry. *Journal of Mathematics Teacher Education, 14*(2), 133-148.

Confrey, J., A. P. Maloney, K. H., Nguyen, G., Mojica, G., & Myers, M. (2009). Equipartitioning / splitting as a foundation of rational number reasoning using learning trajectories. In M. Tzekaki, M. Kaldrimidou, & H. Sakonidis (Eds). *Proceedings of the 33rd Conference of the International Group for the Psychology of Mathematics Education.* Thessaloniki, Greece: PME.

Corbetta, M., & Shulman, G. (2002). Control of goal-directed and stimulus driven attentions in the brain. *Nature Reviews Neuroscience, 3*, 201-215.

Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology, 78*(1) 98-98.

Criswell, B. (2011). Reducing the degrees of freedom in chemistry classroom conversations. *Chemistry Education Research and Practice, 13(*1*)*, 17-29.

Dawson, C., & Wilby, R. (1999). An artificial neural network approach to rainfall runoff modeling. *Hydrological Sciences, 43*(1) 47-66.

Deane, P. The skills underlying writing expertise: Implications for K-12 writing assessment. (2010). *Educational Testing Service Report*. Retrieved from http://144.81.87.152/s/research/pdf/CBALwriting.pdf

DeLong, E. R., DeLong, D. M., & Clarke-Pearson, D. L. (1988). Comparing the areas under two or more correlated receiver operating characteristic curves: A nonparametric approach, *Biometrics, 44*, 837-845.

Demarest, E. J. (2010). *A Learning-centered framework for education reform: What does it mean for national policy?* New York, NY: Teachers College Press.

Demetriadis, S., & Pombortsis, A. (2007). E-lectures for flexible learning: A study on their learning efficiency. *Educational Technology & Society, 10*(2), 147-157.

Demopoulos, W. (2003). On the rational reconstruction of our theoretical knowledge. *The British Journal for the Philosophy of Science, 54*(3), 371-403.

DeYoe, E. A., Bandettini, P., Neitz, J., Miller, D., & Winans, P. (1994). Functional magnetic resonance imaging (FMRI) of the human brain. *Journal of Neuroscience Methods, 54*(2), 171-187.

DeYoung, C. G. (2006). Higher-order factors of the Big Five in a multi-informant sample. *Journal of Personality and Social Psychology, 91*, 1138- 1151.

Dijkstra, T., & Henseler, J. (2011). Linear indices in nonlinear structural equation models; best fitting indices and other composites. *Quality & Quantity, 45*, 1505-1518.

Dimitrov, D. (2003). Marginal true-score measures and reliability for binary items as a function of their IRT parameters. *Applied Psychological Measurement, 27*(6), 440-458.

Dimitrov, D. (2007). Least squares distance method of cognitive validation and analysis for binary items using their item response theory parameters. *Applied Psychological Measurement, 31*, 367-387.

Dimitrov, D. (2008). *Quantitative research in education: Intermediate & advanced methods*. Oceanside, NY: Whittier.

Dimitrov, D. (2012). *Statistical methods for validation of assessment scale data in counseling and related fields.* Alexandria, VA: American Counseling Association.

Dondinger, M. (2007). Educational video game design: A review of the literature. *Computer and Information Science. 4*(1), 21-31.

Efendigil, T., Onut, S., & Kahrman, C. (2009). A decision support system for demand forecasting with artificial neural networks and neuro-fuzzy models: A comparative analysis. *Expert Systems with Application, 36*, 6697-6707.

Elman, B. (2002). *A Cultural History of Civil Examinations in Late Imperial China*. London, England: University of California Press.

Embretson, S., & Gorin, J. (2001). Improving construct validity with cognitive psychology principles. *Journal of Educational Measurement, 38*, 343-368.

ETS website. (2012, November 26). ETS assessment games challenge. Etsgameschallenge.com. Retrieved November 26, 2012, from http://etsgameschallenge.com.

Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods, 44*, 272-299.

Fayol, M., Largy, P., & Lemaire, P. (1994). Cognitive overload and orthographic errors: When cognitive overload enhances subject-verb agreement errors. *The Quarterly Journal of Experimental Psychology*, *47*, 437-464.

Fichtenholtz, H. M., Dean, H. L., Dillon, D. G., Yamasaki, H., McCarthy, G., & LaBar, K. S. (2004). Emotion-attention network interactions during a visual oddball task. *Cognitive Brain Research*, *20*(1), 67-80.

Flaitz, J. (2011). Assessment for learning: US perspectives. In *Assessment Reform in Education* In R. Berry (Ed.), Assessment reform in education (pp. 33-47). Amsterdam, Netherlands: Springer.

Fraenkel, J.R., Wallen, N.E., & Hyun, H.H. (2012). *How to design and evaluate research in education*. 8th ed. New York, NY: McGraw-Hill.

Gallant, S. (1993). *Neural network learning and expert systems.* London, England: MIT Press.

Gauvain, J. L., & Lee, C. H. (1994). Maximum a posteriori estimation for multivariate Gaussian mixture observations of Markov chains. *Speech and Audio Processing, IEEE Transactions on Acoustics*, *2*(2), 291-298.

Gierl, M. (2007). Making diagnostic inferences about cognitive attributes using the rule-space model and attribute hierarchy method. *Journal of Educational Measurement*, *44*(4), 325-340.

Gierl, M., & Leighton, J. (2006*). Simulation studies for evaluating the performance of the two classification methods in the AHM.* Cambridge, MA: Cambridge University Press.

Gierl, M., & Zhou, C. (2009). Reliability and attribute-based scoring of cognitive diagnostic assessment. *Journal of Educational Measurement*, *46*, 293-313.

Glunt, W., Hayden, T. L., Hong, S., & Wells, J. (1990). An alternating projection algorithm for computing the nearest Euclidean distance matrix. *SIAM Journal on Matrix Analysis and Applications*, *11*, 589-600.

Göbel, S. M., & Snowling, M. J. (2010). Number-processing skills in adults with dyslexia. *The Quarterly Journal of Experimental Psychology*, *63*, 1361-1373.

Goddard, J. B., & Kirby, A. (1976). *An introduction to factor analysis*. Norwich, England: Geological Abstracts.

Green, C. D., Feinerer, I., & Burman, J. T. (2013). Beyond the schools of psychology 1: A digital analysis of psychological review, 1894–1903. *Journal of the History of the Behavioral Sciences*, *49*(2), 167-189.

Grefenstette, J. J., Ramsey, C. L., & Schultz, A. C. (1990). Learning sequential decision rules using simulation models and competition. *Machine Learning*, *5*, 355-381.

Gupta, A. (2010). Predictive modeling of turning operations using response surface methodology, artificial neural networks, and support vector regression. *International Journal of Production Research*, *48*, 763-778.

Hadwin, A., Winne, P., & Nesbit, J. (2005). Roles for software technologies in advancing research and theory in educational psychology. *British Journal of Educational Psychology*, *75*(1), 1-24.

Hall, K. (2012). *Grounding assessment in authentic pedagogy: A case study of general education assessment* (Doctoral dissertation). Retrieved from http://ir.stthomas.edu/caps_ed_lead_docdiss/24/

Hambleton, R. K., & Jones, R. W. (1993). Comparison of classical test theory and item response theory and their applications to test development. *Educational Measurement: Issues and Practice*, *12*(3), 38-47.

Hanania, R., & Smith, L. B. (2010). Selective attention and attention switching: Towards a unified developmental approach. *Developmental Science*, *13*, 622-635.

Hanson, J. M., & Mohn, L. (2011). Assessment Trends: A ten-year perspective on the uses of a general education assessment. *Update*, *23*(5), 1-16.

Harvey, R. (1999). Item response theory, *The Counseling Psychologist*, *27*, 353-383.

Hattie, J., & Jaeger, R. (2009). Assessment and classroom learning: A deductive approach. *Assessment in Education*, *5*(1), 111-122.

Hawley, R. J., & Eitzen, Jr, E. M. (2001). Biological weapons-A primer for microbiologists 1. *Annual Reviews in Microbiology*, *55*(1), 235-253.

Hayes, B. K., Heit, E., & Swendsen, H. (2010). Inductive reasoning. *Wiley interdisciplinary reviews, Cognitive science, 1*(2), 278-292.

Hayes, E. & Games, I. (2008). Making computer games and design thinking: A review of current software and strategies. *Games and Culture, 3*(3-4), 309-332.

Hebb, D. (1949). *The organization of behavior: A neuropsychological theory*, Mahwah, NJ: John Wiley & Sons.

Heid, M. K., & Blume, G. W. (2008). Algebra and function development. *Research on Technology and the Teaching and Learning of Mathematics, 1*, 55-108.

Henson, R., Roussos, L., Douglas, J. & He, X. (2008). Cognitive diagnostic attribute-level discrimination indices. *Applied Psychological Measurement, 32*(4), 275-288.

Hestenes, D (2010). Modeling theory for math and science education. In R. Lesh (Ed.), *Modeling Students', Mathematical Modeling Competencies*, 13-41. Amsterdam, Netherlands: Springer.

Hipkins, R. & Kenneally, N. (2003). *Using NEMP to inform the teaching of science skills: Report to the National Education Monitoring Project.* Ministry of Education. Auckland, New Zealand.

Hodson, D. (1985). Philosophy of science, science and science education. *Studies in Science Education, 12*(1), 25-57.

Hoffman, B., & Nadelson, L. (2010). Motivational engagement and video gaming: A mixed methods study. *Educational Technology Research and Development, 58*(3), 245-270.

Hotton, S., & Yoshimi, J. (2011). Extending dynamical systems theory to model embodied cognition. *Cognitive Science, 35*, 444-479.

Hoyer, W. J., Cerella, J., & Buchler, N. G. (2011). A search-by-clusters model of visual search: Fits to data from younger and older adults. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences, 66*, 402-410.

Hubert-Wallander, B., Green, C. S., & Bavelier, D. (2011). Stretching the limits of visual attention: The case of action video games. *Wiley Interdisciplinary Reviews: Cognitive Science, 2*(2), 222-230.

Huff, K., & Goodman, G. (2007). The demand for cognitive diagnostic assessment. In J. Leighton & M. Gierl (Eds), *Cognitive diagnostic assessment: Theory and applications.* (p. 19-60) Cambridge, England: Cambridge University Press.

Hyvärinen, A., & Oja, E. (2000). Independent component analysis: algorithms and applications. *Neural Networks*, *13*(4), 411-430.

Im, S., & Yin, Y. (2009). Diagnosing skills of statistical hypothesis testing using the Rule Space Method. *Studies in Educational Evaluation*, *35*(4), 193-199.

Jamaludin, A., & Chee, Y., & Mei, Lin Ho, C. (2009). Fostering argumentative knowledge construction through enactive role-play in Second Life. *Computers & Education*, *53*(2), 317-329.

Johnson, L., Rickel, J. & Lester, J. Animated pedagogical agents: Face to face interaction in interactive learning environments. *International Journal of Artificial Intelligence in Education,* 11, 47-78.

John-Steiner, V., & Mahn, H. (1996). Sociocultural approaches to learning and development: A Vygotskian framework. *Educational Psychologist*, *31*(3-4), 191-206.

Kalyuga, S., Rikers, R., & Paas, F. (2012). Educational implications of expertise reversal effects in learning and performance of complex cognitive and sensorimotor skills. *Educational Psychology Review*, *24*(2), 313-337.

Kaufman, S. B. (2011). *Intelligence and the cognitive unconscious*. In R. Sternberg & S. Kaufman (Eds.). *The Cambridge handbook of intelligence*, (pp. 442-467). New York, NY: Cambridge University Press.

Kenny, R. F., & McDaniel, R. (2011). The role teachers' expectations and value assessments of video games play in their adopting and integrating them into their classrooms. *British Journal of Educational Technology*, *42*(2), 197-213.

Kim, B., Park, H., & Baek, Y. (2009). Not just fun, but serious strategies: Using meta-cognitive strategies in game based learning. *Computers and Education*, *52*, 800-810.

Kirriemuir, J., & Mcfarlane, A. (2004). Literature review in games and learning (Futurelab Series Report 8). Retrieved from TeLearn website: http://telearn.archives-ouvertes.fr/hal-00190453/

Kitsantas, A., & Zimmerman, B. J. (2009). College students' homework and academic achievement: The mediating role of self-regulatory beliefs. *Metacognition and Learning, 4*(2), 97-110.

Kolkman, M. E., Hoijtink, H. J., Kroesbergen, E. H., & Leseman, P. P. (2013). The role of executive functions in numerical magnitude skills. *Learning and Individual Differences, 24*, 145-151.

Kravitz, D. J., Saleem, K. S., Baker, C. I., & Mishkin, M. (2011). A new neural framework for visuospatial processing. *Nature Reviews Neuroscience, 12*(4), 217-230.

Kroeger, L. A., Brown, R. D., & O'Brien, B. A. (2012). Connecting neuroscience, cognitive, and educational theories and research to practice: A review of mathematics intervention programs. *Early Education & Development, 23*(1), 37-58.

Ku, S. P., Tolias, A. S., Logothetis, N. K., & Goense, J. (2011). fMRI of the face-processing network in the ventral temporal lobe of awake and anesthetized macaques. *Neuron, 70*(2), 352-362.

Kuhn, D. & Matson, J. (2002) A validity study of the screening tool of feeding problems (STEP). *Journal of Intellectual and Developmental Disability, 27*(3), 161-167.

Lai, H., Gierl, M., & Cui, Y. (2012). *Item consistency: An item-fit index for cognitive diagnostic assessment.* Presented at the Annual Meeting for the National Council of Measurement in Education, Vancouver, BC.

Lamb, R., & Annetta, L. (2009). A pilot study of online simulations and problem based learning in a chemistry classroom. *Journal of Virginia Science Education, 3*(2), 34-50.

Lamb, R., & Annetta, L. (2010). Influences of gender on computer simulation outcomes. *Meridian: A Middle School Computer Technologies Journal, 13*(1): http://www.ncsu.edu/meridian/winter2010/lamb/index.htm.

Lamb, R., Annetta, L. Meldrum, J., & Vallett, D. (2011). Measuring science interest: Rasch validation of the science interest survey. *International Journal of Science and Mathematics Education, 10*(3), 643-668.

Lamb R., and Annetta L. (2012). The use of online modules and the effect on student outcomes in a high school chemistry class, *Journal of Science Education and Technology*, Online publication. doi: 10.1007/s10956-012-9417-5

Langley, P., Laird, J. E., & Rogers, S. (2009). Cognitive architectures: Research issues and challenges. *Cognitive Systems Research*, *10*(2), 141-160.

Lechner, M. Lollivier, S. & Magnac, T. (2008). Parametric binary choice models, the econometrics of panel data. *Advanced Studies in Theoretical and Applied Econometrics, 46*(1), 215-245.

Leighton, J. & Gierl, M. (2007). Defining and evaluating models of cognition used in educational measurement to make inferences about examinees' thinking processes. *Educational Measurement: Issues and Practice*, *26*(2), 3-16.

Leighton, J., Gierl, M., Hunka, S. (2006). The attribute hierarchy method for cognitive assessment: A variation on Tatsuoka's rule-space approach. *Journal of Educational Measurement*, *41*(3), 205-237.

Li, R. (2009). *The Effects of Action Video Game Playing on Low Level Vision* (Doctoral dissertation). University of Rochester, NY.

Limniou, M., Roberts, D., & Papadopoulos, N. (2008). Full immersive virtual environment CAVE in chemistry education. *Computers & Education, 51,* 584-593.

Linacre, J. (1991, April). *Structured rating scales*. Paper presented at the Sixth International Objective Measurement Workshop, Chicago, IL.

Linacre, J. (1999). Investigating rating scale category utility. *Journal of Outcome Measurement*, *3*(2), 103-122.

Linhares, A., Freitas, A. E. T., Mendes, A., & Silva, J. S. (2012). Entanglement of perception and reasoning in the combinatorial game of chess: Differential errors of strategic reconstruction. *Cognitive Systems Research*, *13*(1), 72-86.

Lipinski, J., Sandamirskaya, Y., & Schoner, G. (2009). Swing it to the left, swing it to the right: Enacting flexible spatial language using a neurodynamic framework. *Cognitive Neurodynamics*, *3*, 373-400.

Loncella, E. (2010). Integrating computation and visualization to enhance learning IR spectroscopy in general chemistry laboratory: Computer-assisted learning of IR-spectroscopy. *Third Special Issue on Undergraduate Research and Education in Spectroscopy*, *43*(7-8), 618-625.

Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Mahwah, NJ: Routledge.

Lord, F., & Novick, M. (1968). *Statistical theories of mental test scores with contributions from Alan Birnbaum*. Reading, MA: Addison-Wesley.

Maher, J. (2011). *Towards an appreciation of the place and potential of computer games in education*, (Doctoral dissertation). Retrieved from University of Limerick Institutional Repository. (10344 1929)

Mance, I., & Vogel, E. K. (2013). Visual working memory. *Wiley Interdisciplinary Reviews: Cognitive Science*, *4*(2), 179-190.

Mandinach, E. B., Rivas, L., Light, D., Heinze, C., & Honey, M. (2006, April). *The impact of data-driven decision making tools on educational practice: A systems analysis of six school districts.* Paper presented at the Annual meeting of the American Educational Research Association, San Francisco, CA. Retrieved from *http://cct-dev.edc.org/userfiles/publications/speeches/Data_AERA06.pdf*

Manktedlow, K. (2012). *Thinking and reasoning: An introduction to the psychology of reason, judgment, and decision making*. New York, NY: Psychology Press.

Mayer, J. S., Roebroeck, A., Maurer, K., & Linden, D. E. (2010). Specialization in the default mode: Task-induced brain deactivations dissociate between visual working memory and attention. *Human Brain Mapping*, *31*(1), 126-139.

Medin, D. L., & Shoben, E. J. (1988). Context and structure in conceptual combination. *Cognitive Psychology*, *20*(2), 158-190.

Meila, M., & Heckerman, D. (2001). An experimental comparison of model-based clustering methods. *Machine learning*, *42*(1), 1-2.

Messick, S. (1984). The psychology of educational measurement. *Journal of Educational Measurement*, *21*(3), 215-237.

Mitchell, A. & Savill-Smith, 2004, *The use of computer and video games for learning: A review of the literature*. Retrieved from Learning and Skill Development Laboratory website: http://dera.ioe.ac.uk/5270/1/041529.pdf

Moeller, K., Wood, G., Doppelmayr, M., & Nuerk, H. C. (2010). Oscillatory EEG correlates of an implicit activation of multiplication facts in the number bisection task. *Brain Research*, *13*(20), 85-94.

Moreau, D. (2012). The role of motor processes in three-dimensional mental rotation: Shaping cognitive processing via sensorimotor experience. *Learning and Individual Differences*, *22*(3), 354-359.

Moreno-Ger, P. Burgos, D., Martinez-Ortiz, I., Sierra, L., Fernandez-Manjon, B. (2008). Educational game design for online education. *Computers in Human Behavior, 24*, 2530-2540.

National Education Association. (1918). *Cardinal principles of secondary education: A report of the commission on the reorganization of secondary education.* (U.S. Bureau of Education Bulletin No. 35). Washington, D.C.: U.S. Government Printing Office.

National Education Association. (1920). *Reorganization of science in secondary schools: A report of the commission on the reorganization of secondary education.* (U.S. Bureau of Education Bulletin No. 20). Washington, D.C.: U.S. Government Printing Office.

Neary, D. (1997). The prefrontal cortex. *Brain a Journal of Neurobiology, 122*(2), 370a-370j.

Nelder, J. A., & Wedderburn, R. W. (1972). Generalized linear models. *Journal of the Royal Statistical Society. Series A (General), 135*, (Part 3), 370-384.

Ngai, E., Hu, Y., Chen, Y., & Sun, X, (2011). The application of data mining techniques in financial fraud detection: A classification framework and an academic review of the literature. *Decision Support Systems, 50*, 559-569.

Oczkowski, W. J., & Barreca, S. (1997). Neural network modeling accurately predicts the functional outcome of stroke survivors with moderate disabilities. *Archives of Physical Medicine and Rehabilitation, 78*(4), 340-345.

Odena, O. (2010). Practitioners' views on cross-community music education projects in Northern Ireland: Alienation, socio-economic factors and educational potential. *British Educational Research Journal, 36*(1), 83-105.

O'Hara, M. (2012). *Closing the expectations gap, 2012: 50-state progress report on the alignment of K-12 policies and practice with the demands of college and careers.* Retrieved from Educational Resource Information Center website: http://www.eric.ed.gov/ERICWebPortal/search/detailmini.jsp?_nfpb=true&_&ERICExtSearch_SearchValue_0=ED535986&ERICExtSearch_SearchType_0=no&accno=ED535986

Orr, M. (1995). Regularization in the selection of radial basis function centers. *Neural Computation, 7*, 606-623.

Ozuru, Y., Dempsey, K., & McNamara, D. S. (2009). Prior knowledge, reading skill, and text cohesion in the comprehension of science texts. *Learning and Instruction, 19*(3), 228-242.

Papageorgiou, E. I. (2011, June). *Review study on fuzzy cognitive maps and their applications during the last decade.* Paper presented at the *IEEE International Conference*, Taipei, Taiwan, Abstract retrieved from http://ieeexplore.ieee.org/xpl/login.jsp?tp=&arnumber=6007670&url=http%3A%2F%2Fieeexplore.ieee.org%2Fxpls%2Fabs_all.jsp%3Farnumber%3D6007670

Papastergiou, M. (2009). Digital Game-Based Learning in high school Computer Science education: Impact on educational effectiveness and student motivation. *Computers & Education, 52*(1), 1-12.

Parzen, E. (1962). On estimation of a probability density function and mode. *The Annals of Mathematical Statistic*s, *33*, 1065-1076.

Pasqualotto, A., & Proulx, M. J. (2012). The role of visual experience for the neural basis of spatial cognition. *Neuroscience & Biobehavioral Reviews, 36*, 1179-1187.

Peifors, A., Tenenbaum, J. B., Griffiths, T. L., & Xu, F. (2011). A tutorial introduction to Bayesian models of cognitive development. *Cognition, 120*(3), 302-321.

Pennick, M. R., & Kana, R. K. (2012). Specialization and integration of brain responses to object recognition and location detection. *Brain and Behavior, 2*(1), 6-14.

Penuel, W., Fishman, B. J., Gallagher, L. P., Korbak, C., & Lopez-Prado, B. (2009). Is alignment enough? Investigating the effects of state policies and professional development on science curriculum implementation. *Science Education, 93*(4), 656-677.

Perlovsky, L. (2009). Language and cognition. *Neural Networks, 22*(3) 247-257.

Petridou, A., & Williams, J. (2010). The extent of mismeasurement for aberrant examinees. *Educational Assessment, 15*(1), 42-68.

Phillips, V., & Bond, C. (2004). Undergraduates' experiences of critical thinking. *Higher Education Research & Development*, *23*(3), 277-294.

Pinkus, A. (1999). Approximation theory of the MLP model in neural networks. *Acta Numerica*, *8*, 143-195.

Portney L. & Watkins M.(2000). *Foundations of clinical research: Applications to research*. 2nd ed. Upper Saddle, NJ: Prentice Hall Health.

Posner, M. I., & Petersen, S. E. (1989). *The attention system of the human brain* (Report No. TR-89-1). Retrieved from http://oai.dtic.mil/oai/oai?verb=getRecord&metadataPrefix=html&identifier=ADA206157

Pruzek, R. (2005*).* Factor analysis: exploratory. Encyclopedia of statistics in behavioral science. *Journal of the American Statistical Association, 103*(482), 881-882.

Rao, C. R. (1964). The use and interpretation of principal component analysis in applied research. *Sankhyā: The Indian Journal of Statistics, 26*(4), 329-358.

Rasch, G. (1960). *Studies in mathematical psychology: I. Probabilistic models for some intelligence and attainment tests* [Monograph]. Oxford, England: Nielsen & Lydiche.

Rastle, K. (2012). Rethinking phonological theories of reading. *Behavioral and Brain Sciences*, *1*(1), 41-42.

Raykov, T. (2009). Evaluation of scale reliability for unidimensional measures using latent variable modeling. *Measurement and Evaluation in Counseling and Development, 42*(3), 223-232.

Raykov, T., Dimitrov, D. & Asparouhov, T. (2010). Evaluation of scale reliability with binary measures using latent variable modeling, *Structural Equation Modeling: A Multidisciplinary Journal*, *17*(2), 265-279.

Ren, X., Schweizer, K., & Xu, F. (2013). The sources of the relationship between sustained attention and reasoning. *Intelligence, 41*(1), 51-58.

Reynolds, S., Lane, S. J., & Richards, L. (2010). Using animal models of enriched environments to inform research on sensory integration intervention for the

rehabilitation of neurodevelopmental disorders. *Journal of Neurodevelopmental Disorders*, *2*(3), 120-132.

Rezaee, A. A., & Azizi, Z. (2012). The role of zone of proximal development in the students' learning of English adverbs. *Journal of Language Teaching and Research*, *3*(1), 51-57.

Riding, R., & Cheema, I. (1991). Cognitive styles—an overview and integration. *Educational Psychology*, *11*(3-4), 193-215.

Roberts, M., & Gierl, M. (2010). Developing score reports for cognitive diagnostic assessments. *Educational Measurement: Issues and Practice*, *29*(3), 25-38.

Sarama, J., & Clements, J. (2002). Building blocks for young children's mathematical development. *Journal of Educational Computing Research, 27*(1), 93-110.

Schmitz, N., Rubia, K., van Amelsvoort, T., Daly, E., Smith, A., & Murphy, D. (2008). Neural correlates of reward in autism. *The British Journal of Psychiatry*, 192, (19-25).

Sharma, S. K., & Kitchens, F. L. (2004). Web services architecture for m-learning. *Electronic Journal on e-Learning*, *2*(1), 203-216.

Smith, R. M. (1991). The distributional properties of Rasch item fit statistics. *Educational and Psychological Measurement*, *51*, 541-565.

Smith, K., & Gupta, J. (2003). *Neural networks in business: techniques and applications* for the operations researcher. *Computers & Operations Research, 27*, 1023-1044.

Soares, T. (2009). An integrated Bayesian Model for DIF Analysis. *Journal of Educational and Behavioral Statistics*, *34*(3), 348-377.

Somoza, E., & Somoza, J. R. (1993). A neural-network approach to predicting admission decisions in a psychiatric emergency room. *Medical Decision Making*, *13*(4), 273-280.

Songer, N., Kelcey, B., & Gotwals, A. (2009). How and when does complex reasoning occur? Empirically driven development of a learning progression focused on complex reasoning about biodiversity. *Journal of Research in Science Teaching*, *46*(6), 610-631.

Spada, H. (1977). *Logistic models of learning and thought, Structural models of thinking and learning.* Bern, Switzerland: Huber.

Specht, D. F. (1991). A general regression neural network. *IEEE Transactions on Neural Networks*, *2*(6), 568-576.

Spector, M., & Changmin, K. (2012). A model-based approach for assessment and motivation. *Computer Science and Information Systems*, *9*(2), 893-915.

Spence, I., & Feng, J. (2010). Video games and spatial cognition. *Review of General Psychology*, *14*(2), 92-104.

Song, H. J. (2011). *Evaluation of the effects of spatial separation and timbral differences on the identifiability of features of concurrent auditory streams* (Doctoral dissertation). Retrieved from The Sydney eScholarship Repository. (2123 7213)

Squire, K. D. (2003). Video games in education. *International Journal of Intelligence: Games & Simulation*, *2*(1), 49-62.

Steen, L. (2010). The science of patterns, *Science*, *240*, 611-616.

Sternberg, R.J. (Ed.)(1982). *Handbook of human intelligence*, New York, NY: Cambridge University Press.

Stewart, D. W. (1981). The application and misapplication of factor analysis in marketing research. *Journal of Marketing Research*, *18*(1), 51-62.

Suppes, P. (1969). Probabilistic grammars for natural languages, *Synthese*, *22*(1-2), 95-116.

Sventina, D., Gorin, J. S., & Tatsuoka, K. K. (2011). Defining and comparing the reading comprehension construct: A cognitive-psychometric modeling approach. *International Journal of Testing*, *11*(1), 1-23.

Tatsuoka, K. (1983). Rule space: An approach for dealing with misconceptions based on item response theory. *Journal of Educational Measurement*, *20*(4), 345-354.

Tesio, L. (2003). Measuring behaviours and perceptions: Rasch analysis as a tool for rehabilitation research. *Journal of Rehabilitation Medicine*, *35*(3), 105-115.

Teufel, C., Fletcher, P. C., & Davis, G. (2010). Seeing other minds: attributed mental states influence perception. *Trends in Cognitive Sciences*, *14*(8), 376-382.

Thomas, P., & Macredie, R. (1994). Games and the design of human-computer interfaces. *Innovations in Education and Training International*, *31*(2), 134-142.

Thurston, L. (1947). *Multiple factor analysis.* Chicago, IL: University of Chicago Press

Tolentino, L., Birchfield, D., Megowan-Romanowicz, C., Johnson-Glenberg, M., Kelliher, A., & Martinez, C. (2009). *Journal of Science Education and Technology, 18*(6), 501-517.

Trilling, B., & Fadel, C. (2009). *21st century skills: learning for life in our times*. Retrieved from http://www.21stcenturyskillsbook.com/index.php

U.S. Department of Education. National Center for Education Statistics (1998). *Pursuing Excellence: A Study of U.S. Twelfth-Grade Mathematics and Science Achievement in International Context* (NCES 98-049). Washington, DC: U.S. Government Printing Office.

Uttal, D., & Cohen, C. (2012). Spatial thinking and STEM education, when, why and how? In B. Ross (Ed.), *The psychology of learning and motivation* (pp. 147-178). Amsterdam, Netherlands: Academic Press.

Volet, S., Vauras, M., & Salonen, P. (2009). Self and social regulation in learning contexts: An integrative perspective. *Educational Psychologist*, *44*(4), 215-226.

Von Daver, M. (2010). A general diagnostic model applied to language testing data. *British Journal of Mathematical and Statistical Psychology, 61*(2), 287-307.

Vygotsky, L. (1978). *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.

Waldmann, M. R. (2012). Predictive versus diagnostic causal learning: Evidence from an overshadowing paradigm. *Psychonomic Bulletin & Review, 8*(3), 600-608.

Walls, L. (2011). Third grade African American students views of the nature of science. *Journal of Research in Science Teaching, 49*(1), 1-37.

Wallander, L. (2009). 25 years of factorial surveys in sociology: A review. *Social Science Research, 38*, 505-520.

Wang, C., & Barrow, L. (2011). Characteristics and levels of sophistication: An analysis of Students' abilities to think with mental models. *Research in Science Education, 41*, 561-586.

Wang, C., & Chang, H. (2008). Item selection in multidimensional computerized adaptive testing-gaining information from different angles. *Psychometrika, 76*(3), 363-384.

Wang, Y., Geng, F., Yuzheng, H., Du, F., & Chen, F. (2013). Numerical processing efficiency improved in experience mental abacus children. *Cognition*, *127*(2), 149-158.

Wang, J., Jackson, L. & Zhang, D. (2011). The mediator role of self-disclosure and moderator roles of gender and social anxiety in the relationship between Chinese adolescents' online communication and their real-world social relationships. *Computer in Human Behavior*, *27*(6), 2161-2168.

Wang, C., Chang, H., & Douglas, J. (2012). The linear transformation model with frailties for the analysis of item response times. *British Journal of Mathematical and Statistical Psychology*, *66*(1), 144-168. doi: 10.1111/j.2044-8317.2012.02053

Wilson, R. A., & Keilm, F. C. (2001). *The MIT encyclopedia of cognitive science*. Cambridge, MA: MIT Press.

Winkler, L. Y. (2009*). Convergence of pitch and number word magnitude coding in the intraparietal cortex* (Doctoral dissertation), Universitätsbibliothek, Berlin, Germany.

Woelert, P. (2012). Idealization and external symbolic storage: the epistemic and technical dimensions of theoretic cognition. *Phenomenology and the Cognitive Sciences, 11*(3), 335-366.

Wu, H. & Shah, P. (2004). Exploring visuospatial thinking in chemistry learning. *Science Education*, *88*(3), 465–492.

Wu, M.L., & Richards, K. (2012). Learning with educational games for the intrepid 21st Century learners. P. Resta (Ed.), *Proceedings of Society for Information Technology & Teacher Education International Conference 2012* (pp. 55-74). Chesapeake, VA: AACE.

Xu, Y., Meyer, K., & Morgan, D. (2008). Piloting a blended approach to teaching statistics in a college of education: Lessons learned. *The Journal of Educators Online, 5*(2), 1-20.

Yang, X., Tong, J., Xu, X. (2009). Visual effects in computer games. *Computer, 42*(7), 48-56.

Yang, B. (2005). *Factor analysis methods.* In R. Swanson & E. Holton (Eds.), (pp.181-199). San Francisco, CA: Berrett-Koehler Publishers.

Young, M. B. (2011). *One journey, several destinations: an exploratory study of local contextualization of national assessment policy* (Doctoral dissertation). Retrieved from DART-Europe E-theses Portal. (2649)

BIOGRAPHY

Richard L. Lamb was born on September 13, 1976 in New York. He attended and graduated from Sweet Home Senior High School, Amherst New York in 1994. He received his Bachelor of Science from Canisius College in 1999. Upon graduation, he entered the United States Army as a Second Lieutenant. Upon exiting the Army, he was employed as a teacher in North Carolina and Virginia for a total of eight years and received his Master of Science in Science Education under the direction of Leonard A. Annetta, Ph.D. Committee Chair, John C. Park, Ph.D., and Michael M. Kimberley at North Carolina State University in 2008.