12-2010

The Multiple-choice concept map (MCCM): An interactive computer-based assessment method

Ioan C. Sas
University of Nevada, Las Vegas

Follow this and additional works at: http://digitalscholarship.unlv.edu/thesesdissertations

Part of the Cognitive Psychology Commons, Educational Assessment, Evaluation, and Research Commons, and the Educational Psychology Commons

Repository Citation
http://digitalscholarship.unlv.edu/thesesdissertations/769

This Dissertation is brought to you for free and open access by Digital Scholarship@UNLV. It has been accepted for inclusion in UNLV Theses, Dissertations, Professional Papers, and Capstones by an authorized administrator of Digital Scholarship@UNLV. For more information, please contact digitalscholarship@unlv.edu.
THE MULTIPLE-CHOICE CONCEPT MAP (MCCM):
AN INTERACTIVE COMPUTER-BASED
ASSESSMENT METHOD

by

Ioan Ciprian Sas

Bachelor of Theology
St. Andrew’s College, Winnipeg, Manitoba
1998

Master of International Business
Saint Mary’s University of Minnesota
1999

A dissertation submitted in partial fulfillment of
the requirements for the

Doctor of Philosophy in Learning and Technology
Department of Educational Psychology
College of Education

Graduate College
University of Nevada, Las Vegas
December 2010
THE GRADUATE COLLEGE

We recommend the dissertation prepared under our supervision by

Ioan Ciprian Sas

entitled

The Multiple-Choice Concept Map (MCCM): An Interactive Computer-Based Assessment Method

be accepted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Learning and Technology
Department of Educational Psychology

Kendall Hartley, Committee Chair

Michael Nussbaum, Committee Member

Kent J. Crippen, Committee Member

Sterling J. Saddler, Graduate Faculty Representative

Ronald Smith, Ph. D., Vice President for Research and Graduate Studies
and Dean of the Graduate College

December 2010
This research attempted to bridge the gap between cognitive psychology and educational measurement (Mislevy, 2008; Leighton & Gierl, 2007; Nichols, 1994; Messick, 1989; Snow & Lohman, 1989) by using cognitive theories from working memory (Baddeley, 1986; Miyake & Shah, 1999; Grimley & Banner, 2008), multimedia learning (Mayer, 2001), and cognitive load (Chandler and Sweller, 1991, 1992; Cerpa, Chandler, & Sweller, 1996) to identify potential design weaknesses of traditional select-and-fill-in (SAFI) concept map assessment and then to guide the design of the new and improved multiple-choice concept map (MCCM) assessment method.

The primary objective of this study was to evaluate the effects of (1) the type of the list of concepts or relations and (2) the spatial placement of the selection list on the examinee’s overall mental effort. Using a 2x2 Factorial MANCOVA, with prior knowledge as the covariate, the participants were compared on the length of time to complete the assessment and on examinee’s rating of personal mental effort exerted during the assessment. A Simple Planned Comparison was conducted to evaluate estimated mean differences between the primary group of interest (MCCM-integrated
Group 4) and each of the other three groups. Of additional interest were differences among the four different groups on the examinees’ attitudes and impressions toward their respective assessment method; these data provided more insight into the quality of design for each treatment group.

The type of List/Map had a significant main effect on the overall mental effort/strain or the linear combination of time on task and mental effort rating scores. More specifically, however, the test of between-subjects effects indicated that the type of list/map had a statistically significant effect on the time to complete the assessment, but not on the examinees’ mental effort ratings. In other words, participants who received the MCCM assessments (which used the MCL) completed the assessment task significantly faster but did not rate higher or lower on personal mental effort than the participants who received the traditional SAFI concept map assessments.

The spatial placement of the selection list (integrated and non-integrated) did not have a statistically significant main effect on the linear combination of time on task and mental effort rating scores. The test of between-subjects effects indicated that the spatial placement of the selection list (integrated versus non-integrated) did not have a statistically significant effect on the time to complete the assessment or on the examinees’ exerted mental effort. In other words, participants who received the integrated lists did not complete the assessment task faster and did not rate themselves lower on personal mental effort than the participants who received non-integrated selection list.

The estimated mean time on task of the primary group of interest (MCCM-integrated Group 4) was significantly shorter than that of both SAFI groups (integrated
and non-integrated), but was not significantly shorter when compared to the time on task of the MCCM-non-integrated Group 3. There were no significant differences on mental effort ratings between MCCM-integrated Group 4 and either of the other three groups.

The participant responses to the Attitudes and Impressions Questionnaire indicated a general preference for the MCCM over the more traditional SAFI map and, in some cases, a slight preference of MCCM-integrated over MCCM-non-integrated.

The multiple-choice concept map (MCCM) is a new interactive computer-based assessment method that bridges well-documented research findings from cognitive science and psychometrics and has the potential to be an external knowledge representation with much-desired characteristics of being valid, efficient, and explicit (Mislevy, Behrens, Bennett, DeMark, Frezzo, Levy, Robinson, Rutstein, Stanley, Winters, & Shute, 2007) when assessing students’ knowledge structures or schema-based knowledge (Pellegrino, Chudowsky, & Glasser, editors, 2001; NAEP, 2008).
ACKNOWLEDGEMENTS

“For with God nothing will be impossible” (Luke 1:37) – These powerful words uttered by Archangel Gabriel to the Virgin Mary summarize my sentiment and firm belief about the completion of this Dissertation and my doctoral degree. In the midst of unpredictable, uncontrollable, and sometimes very unfavorable circumstances, it is indeed only faith, hope, and love that pull us through and, in the end, render us victorious. Humbled by the entire experience of completing all doctoral degree requirements, I offer my first thanks and appreciation to the one who needs no acknowledgement – God. In spite of numerous unfavorable conditions that caused so many tasks to seem impossible, God made this great accomplishment possible for me. Glory be to Him!

I would like to acknowledge my beloved wife, Dr. Magdalena Sas, and our children: Elijah, Emanuela, and Isaiah. Dear Maggie, your direct inputs and recurrent sacrifices during the completion of these degrees (yours and mine) are monumental and truly expressive of your love for me and our family. We met on Orientation Day (August 20, 2001) and our only task then was to complete this doctoral degree, but along the way… we got to know each other; we were married; I was ordained to the priesthood and immediately assumed priestly duties and responsibilities; we were blessed with Elijah and then Emanuela; we moved to Iowa where I was appointed a full time priest with a very unpredictable and irregular schedule; we faced the Floods of 2008 and some health issues; and then we were blessed with our third-born Isaiah. Thus, completing our degrees under these circumstances was no small task and I want you to know that there is no one else I could have or would have rather done this with, but you. I married you for
love and that is what we have. Your intellect (especially your statistical expertise) is a bonus. Thanks for all your help and encouragement. My dearest children (Elijah, Emanuela, and Isaiah), each of your births brought more light and meaning to my life. Your adult-like understanding when daddy had to work on his big book (dissertation) is much appreciated but I want you to know that having to dedicate much of my little spare time to work on the dissertation instead of playing with you was the hardest choice daddy ever had to make. Now that it’s all done, I’m ready to play.

My dear parents, V. Rev. Fr. Aurel Sas and Preoteasa Lidia Sas, your contributions toward all of my accomplishments, including this dissertation, are very significant and I thank you. Dear taticu (dad), you are without a doubt the most disciplined and self-motivated man I know. Had I functioned on the same level and taken all your advice, I would have surely finished this degree much sooner. Thank you for all your prayers and your steadfast encouragement. Dear mamica (mom), I perceive you to be the mother who is more in tune with my feelings than even I am at times. I know that your thoughts and worries about how my efforts toward this dissertation would affect my health, my family, and my ministry far exceeded my own. Thank you for your subtle but very effective encouragement and your perpetual prayers. My dearest parents, I love you both. Your willful sacrifices to bring my siblings and I to the free world and your unceasing efforts throughout your lives as our parents are, in my humble opinion, unequalled. Thus, in appreciation for all that you have done for me, I dedicate this dissertation to both of you.

In addition to my unbelievably sacrificial and loving parents, I also have 4 siblings, the likes of which is just as scarce. I am proud to have initiated life (spiritual,
personal, and professional pursuits alike) among the four of you. I know I would be less
today without all of you in my life, just as we all agree we would be nothing without Him
in our lives. Collectively, you are for me one source of strength and stable reminder of
who I am and who I aspire to be. And now that we all have our life companions, we bond
together into a bigger, stronger knot. Thank you to all my siblings, cousins, as well as
their spouses and children for your prayers and your encouragement during these most
recent academic pursuits.

In addition to being a part of my immediate and extended family, I have also
become the pastor of the wonderful parish family of St. John the Baptist Greek Orthodox
Church in Cedar Rapids, IA. I identify with this tightly knit group of devoted parishioners
very well and feel their love during my ministry. Your encouragement and prayers (in
some cases, very persistent prayers), have been another source of strength for me,
especially during those more pressing conditions. Of the many names I would like to
mention, I will have to restrict myself to three different couples who have opened their
homes to me so as to provide the much needed professional retreats free of all
distractions. Thank you Ken and Edie Yerington, Darrell and Goldie Helgens, and Brandt
and Sandy Worley. Your charming and quiet surroundings enabled me to accomplish in
several days more than I would in months when not in your homes. Each of your homes
has become for me another home-away-from-home. Thank you!

Among the many who have been supportive during this long journey of doctoral
studies are four hierarchs (Metropolitan Iakovos, Bishop Demetrios, Archbishop Nicolae,
Bishop Ioan Casian) and many fellow priests and their wives whom I know have prayed
for my success. Thank you all for your prayers.
Last but certainly not least, I would like to acknowledge my advisor, Dr. Kendall Hartley, and my doctoral committee members: Dr. E. Michael Nussbaum, Dr. Kent Crippen, and Dr. Sterling Saddler. These academically critical minds behind my doctoral research and dissertation writing have been major contributors to my work. Their patience and understanding during my pressing circumstances is also much appreciated. Thank you all for serving on my committee.

Sincerely,

Ioan (John) Ciprian Sas
# TABLE OF CONTENTS

**ABSTRACT** ....................................................................................................................... iii

**ACKNOWLEDGEMENTS**........................................................................................................ vi

**TABLE OF CONTENTS**.................................................................................................... x

**LIST OF TABLES**........................................................................................................... xiii

**LIST OF FIGURES**......................................................................................................... xiv

**CHAPTER 1** INTRODUCTION ..................................................................................... 1  
Purpose of the Study ....................................................................................................... 6  
Issue with the Quality of the Selection List ............................................................... 6  
Improving the Quality of the Selection List ............................................................... 7  
Issue with the Placement of the Selection List ........................................................... 9  
Improving the Placement of the Selection List ......................................................... 10  
Significance of Study .................................................................................................... 11  
Research Questions and Hypotheses ................................................................. 13  
Research Question 1 and Hypothesis 1 ................................................................. 14  
Research Question 2 and Hypothesis 2 ................................................................. 15  
Research Question 3 and Hypothesis 3 ................................................................. 15  
Research Question 4 and Hypothesis 4 ................................................................. 15  
Definition of Terms....................................................................................................... 16

**CHAPTER 2** REVIEW OF LITERATURE .................................................................. 19  
The Development of Educational Assessment in the USA ........................................... 20  
Historical Developmental Milestones Lead to Excessive Use of Standardized Testing and of Multiple-Choice Items ................................................................. 20  
The Rationale behind Alternative Assessments ......................................................... 26  
Alternative Assessments ........................................................................................... 31  
Concept Maps as an Alternative Assessment Method .................................................. 33  
Defining Concept Maps ............................................................................................ 34  
Overcoming Initial Time and Cost Inefficiencies ..................................................... 35  
Validity and Reliability ............................................................................................. 38  
Comparing Concept Map Assessment to Other Types of Assessments ................. 40  
Comparing Different Concept Map Assessment Methods or Tasks ....................... 44  
The Need to Bridge the Gap Between Cognitive Science and Educational Assessment .......................................................... 49  
Cognitive Implications for Traditional SAFI Concept Maps ...................................... 51  
The Traditional Concept Map Selection List ............................................................ 51  
The Spatial Placement of the Concept Map Selection List ..................................... 53  
Cognitive Science and Psychometrics Inform Educational Assessment Design ........ 56  
The Multiple-Choice List (MCL) Used by the Multiple-Choice Concept Map (MCCM) ......................................................................................................................... 56
LIST OF TABLES

Table 1    Demographic Information ................................................................. 111
Table 2    Descriptive Statistics ................................................................. 112
Table 3    Pearson Correlations ................................................................. 113
Table 4    2x2 Factorial MANCOVA for the Composite DV of
      Time on Task and Mental Effort ........................................................... 114
Table 5    Between-Subjects Effects on the 2 Dependent Variables (DV)........... 115
Table 6    Simple Planned Comparison ......................................................... 116
Table 7    Quantitative Data – Percentages for Questions 1-7, 11
      from the Attitudes Questionnaire ..................................................... 117
Table 8    Qualitative Data – Percentages for the Open-Ended
      Questions 8, 9, and 10 ..................................................................... 118
## LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 1</td>
<td>Concept map on the topic of correlation</td>
<td>119</td>
</tr>
<tr>
<td>Figure 2</td>
<td>Practice concept map on the topic of cats and dogs</td>
<td>120</td>
</tr>
<tr>
<td>Figure 3</td>
<td>SAFI concept map used by Schau, Mattern, Zeilik, &amp; Teague (2001)</td>
<td>121</td>
</tr>
<tr>
<td>Figure 4</td>
<td>SAFI concept map used by Chang, Sung, &amp; Chen (2001)</td>
<td>122</td>
</tr>
<tr>
<td>Figure 5</td>
<td>The Select-and-Fill-in (SAFI) concept map non-integrated – Group 1</td>
<td>123</td>
</tr>
<tr>
<td>Figure 6</td>
<td>The Multiple-Choice Concept Map (MCCM) non-integrated – Group 3</td>
<td>124</td>
</tr>
<tr>
<td>Figure 7</td>
<td>The Multiple-Choice Concept Map (MCCM) integrated – Group 4</td>
<td>125</td>
</tr>
<tr>
<td>Figure 8</td>
<td>The Select-and-Fill-in (SAFI) concept map integrated – Group 2</td>
<td>126</td>
</tr>
<tr>
<td>Figure 9</td>
<td>Practice MCCM on cats and dogs integrated – Group 4</td>
<td>127</td>
</tr>
<tr>
<td>Figure 10</td>
<td>Practice MCCM on cats and dogs non-integrated – Group 3</td>
<td>128</td>
</tr>
<tr>
<td>Figure 11</td>
<td>Practice SAFI on cats and dogs integrated – Group 2</td>
<td>129</td>
</tr>
<tr>
<td>Figure 12</td>
<td>Practice SAFI on cats and dogs non-integrated – Group 1</td>
<td>130</td>
</tr>
<tr>
<td>Figure 13</td>
<td>The mental effort rating scale</td>
<td>131</td>
</tr>
</tbody>
</table>
CHAPTER 1
INTRODUCTION

Educational assessment in the United States of America has been continuously evolving ever since the establishment of the first Board of Education in 1837. The initial casual testing of one or a few students by one traveling teacher in the early 1800s due to early geographical challenges has developed into complex and sophisticated educational assessment methods backed by psychometric and cognitive theories. Significant changes in educational assessment practices were prompted by U.S. psychologists whose interests were stimulated by the widely acclaimed effectiveness of Alfred Binet’s first standardized intelligence scales of 1905, 1908, 1911 (Altenbaugh, 1999).

U.S. psychologists contributed to the evolution of the intelligence scales by modifying them to eliminate potential subjectivity/bias present at scoring and by adapting them so as to make them into tests that could be administered to large groups of people. According to Kennedy (2003), the American Psychological Association (APA) administered the Army Alpha and Army Beta tests to nearly 1.7 million young men during WWI. These tests’ desirable qualities of (1) effective and efficient administration to large groups of examinees and (2) objective and efficient scoring were mainly attributed to the type of questions it contained – the multiple-choice-single-selection style question, presently known as the ‘multiple-choice item’. Despite relatively short periods of historical and political unease, multiple-choice type testing, especially when used for large-scale assessment, was considered the gold standard for many decades.

Developmental research in the cognitive and measurement sciences during the past 25 years revealed concerns about the validity of the widely-used multiple-choice
assessment. According to Banks (2005), multiple-choice questions often only measure surface or simple declarative knowledge. Messick (1989) states that

the heart of the notion of the so-called content validity is that the test items are samples of a behavioral domain or item universe about which inferences are to be drawn. But these inferences are likely to invoke, even if only tacitly, psychological processes or behaviors rather than mere surface content. (p. 36)

Drawing test-based inferences about students’ mental or psychological processes or behaviors can guide the improvement of instruction. Thus, Messick’s point that we need to modify assessment practices in order to assess that about which we care to infer is well-grounded. In other words, valid test scores are needed to make accurate inferences about students’ cognitive processes and knowledge. Sadly however, even “accuracy [of response on multiple-choice tests] is not viewed as a good indicator of cognitive processing because a correct answer can be generated by testwise strategies or even in the presence of subtle misconceptions” (Gierl, Leighton, & Hunka, 2007, pp. 265-266).

The need for increasing the quality of educational assessment was formally addressed on a large scale in 2001 by the Committee on the Foundations of Assessment, which states that focus must be placed on assessment methods and practices that measure student acquisition of higher level knowledge (Pellegrino, Chudowsky, & Glasser, editors, 2001), also referred to as “coherent structures of knowledge” (p. 23) or “organization of knowledge” (p. 27). Thus, emphasis is placed on alternate forms of assessment that do a better job than traditional multiple-choice tests not only at detecting or revealing how the examinee’s knowledge is organized, but also at measuring the examinee’s ability to make connections and understand the relationships among concepts. Assessing this type of deeper knowledge is strongly supported by early research on expertise (Chi, Feltovich, & Glaser, 1981; Larkin, McDermott, Simon, & Simon, 1980;
Kalyuga, Chandler, & Sweller, 1998), which demonstrated that well-developed knowledge structures or schemas are essential for expert performance.

The need to assess schematic knowledge structures has driven many educational researchers and psychometricians to investigate alternate forms of assessment such as essay writing, open-ended questions, interviews, think-alouds, concept maps, performance assessments, portfolios, and others. Concept maps quickly became a favorite among these alternate forms of assessment primarily due to their ability to assess the type of knowledge mentioned earlier, namely connected understanding, knowledge organization, or schema-based knowledge structures associated with higher level cognitive thinking/processes (Pellegrino et al., 2001; Kinchin, Hay, & Adams, 2000; Ruiz-Primo, Schultz, Li, & Shavelson, 2001; Yin, Vanides, Ruiz-Primo, Ayala, & Shavelson, 2005). In the report Science Framework for the 2009 National Assessment of Educational Progress (NAEP), the National Assessment Governing Board of the US Department of Education confirmed that concept map assessment can “tap a science ability that is difficult to measure by other means” (NAEP, 2008, p. 102) and encouraged the classroom use and further exploratory research of concept map assessment.

According to O’Neil & Klein (1997), concept maps are graphical representations of students’ understanding of interrelated concepts. They contain labeled nodes/bubbles in the shape of circles or rectangles that are connected by labeled links/lines or lined labels (see Figures 1 and 2). The nodes are labeled by concepts and the connecting links/lines are labeled by the relationships among the concepts. In essence, concept maps are a visual rendering or representation of knowledge organization or schema-based knowledge structures (McClure, Sonak, & Suen, 1999; Novak, 2003) and are, therefore,
deemed to be a desirable form of assessment by the National Assessment Governing Board (NAEP, 2008).

Concept maps have been investigated as both classroom and large-scale assessment. The National Center for Research on Evaluation, Standards, and Student Testing (CRESST) in Los Angeles, CA addressed initial objectivity and cost inefficiencies of scoring concept maps. This center developed the CRESST software that scores paper-and-pencil-administered concept maps by comparing them to expert concept maps (O’Neil & Klein, 1997). Subsequent researchers focused on finding effective ways to accurately, objectively, and efficiently score a wide range of concept map assessment methods and strategies (McClure, Sonak, & Suen, 1999; Ruiz-Primo, Schultz, Li, & Shavelson, 2001; Klein, Chung, Osmundson, Herl, & O’Neil, 2002; Yin & Shavelson, 2004, 2008; Clariana, Koul, & Salehi, 2006; & Keppens & Hay, 2008). Increasingly positive results about the validity and reliability of concept map assessment methods or strategies helped boost confidence for the increased use of concept map assessment.

Concept map assessment methods or techniques range from the student/examinee constructing a map from scratch with minimal direction to the student/examinee completing a partially filled teacher/expert-constructed map by selecting from lists of concepts and/or relations. Yin and Shavelson (2004, 2008) suggest that certain types of concept map assessment methods possess more desirable psychometric properties that render those methods more appropriate for summative, and even large-scale, assessment.

For example, Schau, Mattern, Seilik, Teague, and Weber (2001) found strong evidence of internal consistency and concurrent validity of select-and-fill-in (SAFI) concept maps. SAFI concept maps require examinees/students to fill in blank nodes
and/or links (line labels) into a partially filled teacher/expert-constructed concept map. The selection list(s) used to fill in the map are generally presented separate from or adjacent to the actual concept map. Think-aloud protocols and expressed opinions of students and teachers indicated that the SAFI concept mapping method was easy to use and students employing it utilized “strategies requiring connected understanding” (Schau et al., 2001, p. 148).

In addition to the psychometric benefits of SAFI concept maps over student-constructed concept maps, Schau and colleagues (2001) also point out the following advantages of SAFI concept maps: (1) students do not need to learn how to draw concept maps, as this activity may be perceived to be time-consuming, tedious and frustrating, (2) the quality of SAFI concept maps does not depend on the individual’s communication (language/verbal) skills as it does with student-constructed concept maps, and (3) scoring SAFI concept maps is easier, more accurate, more objective and more efficient than scoring student-constructed concept maps, as there is currently no universally accepted and simple scoring system for student-constructed concept maps.

Nevertheless, traditional SAFI concept maps may have their weaknesses in that the assessment design of the traditional SAFI concept maps may likely impose unnecessary strain on the examinee’s limited cognitive capacity. The next section of this chapter introduces two potential assessment design shortcomings of traditional SAFI concept maps and briefly describe this research study examined these theorized limitations.
Purpose of the Study

The purpose of this study was to examine two potential areas of improvement in the assessment design of traditional SAFI concept maps by evaluating the effects of (1) the type of selection lists of concepts and relations and (2) the spatial placement of these selection lists on the examinee’s overall mental effort.

Issue with the Quality of the Selection List

The first area of improvement of traditional SAFI concept maps pertaining to the type of list(s) of concepts and/or relations to be selected-and-filled-in the partially filled expert concept map is made evident by research findings in both cognitive science and psychometrics. The traditional list (TL) of concepts and/or relations used by traditional SAFI concept maps usually contains about 10 to 20 concepts or relations or a combination of both concepts and relations (see Figures 3, 4, and 5). The assessment task of the student/examinee is to select a concept or relation from the list and place/write/type/click it into the appropriate blank node or blank link within the concept map; naturally, this task is replicated with every attempt to fill in a blank node or blank link. Unfortunately, sifting through this list, or these lists, and doing so repeatedly can be time-consuming, frustrating, and even cognitively overwhelming.

The potential cognitive strain mentioned above has direct implications for working memory (Grimley & Banner, 2008; Pickering & Gathercole, 2004; Miyake & Shah, 1999; Baddeley, 1986), which is the main information processor or the locus of control for our cognitive processes and is limited in both capacity and duration when handling information (Baddeley, 1986; Miyake & Shah, 1999). For example, early research on memory supported that we can remember a series of approximately seven
random numbers (Miller, 1956) for a mere several seconds (Peterson & Peterson, 1959), unless the numbers are intentionally rehearsed. Grimley and Banner (2008) documented the negative effects of low working memory on educational achievement on the General Certificate of School Education (GCSE) exam, namely that students with low working memory capacity are at risk of attaining low GCSE scores. If current concept map assessment design imposes additional cognitive strain on the map user, then the test performance of all students (those with low and high working memory capacity) may be impaired and their test scores negatively affected. Further consideration for the limitations of working memory has led researchers to encourage the design of assessments that eliminate content/construct-irrelevant components (Mayer, 2001; Stout, 2002; Baker, 2007; Mislevy, Bejar, Bennett, Haertel, & Winters, 2010).

Processing 10 to 20 pieces of information on a traditional list (TL) used by SAFI concept maps may cognitively overload the student’s limited working memory capacity. When the examinee reviews the selection list of 10 to 20 concepts/relations to fill in a given blank node or link, most of the items on the list are not related to the corresponding blank node or link as they belong to another blank node or link. In addition, the students/examinees have to view and review the same list, or sometimes lists, with every attempt to fill in a blank node or link, thus being forced to process the same pieces of information repetitively; this redundant task inadvertently and unnecessarily solicits invaluable, limited cognitive resources.

**Improving the Quality of the Selection List**

This study addressed the aforementioned weakness of traditional SAFI concept maps by improving the quality of the list(s) of to-be-selected-and-filled-in concepts and
relations according to relevant research findings from cognitive science and psychometrics. Based on research findings from cognitive science dealing with working memory and human cognitive architecture and its limitations (Miller, 1956; Peterson & Peterson, 1959; Baddeley, 1986; Sweller, van Merrienboer, & Paas, 1998; Kalyuga, Ayres, Chandler, & Sweller, 2003), the number of options (in the form of concepts or relations) contained in a list was reduced by excluding all non-related or construct-irrelevant alternatives, so as to avoid unnecessary cognitive processing. Based on psychometric research findings on the optimum number of alternatives/options for multiple-choice questions (Waller, 1989; Hutchinson, 1997; Abad, Olea, & Ponsoda, 2001), the lists included no more than four concepts or relations to choose from. These four carefully-chosen alternatives (whether concepts or relations) contained one single correct answer and three incorrect-but-plausible alternative(s) to the corresponding blank node or blank link; therefore, each blank node and each blank link needed to have its corresponding carefully-tailored multiple-choice type list of four (or fewer) alternatives.

The presentation of these numerous multiple-choice lists (MCLs) nearby or adjacent to the concept map simultaneously, whether on paper or on a computer screen, would be ineffectual as they would visually and cognitively overwhelm the student. However, in the case of the computer-based concept maps of this study, this potential visual clutter was easily avoided through the use of background programming that enabled the student to view only the multiple-choice list (MCL) that corresponded to the blank node or blank link that the student chose by a single (left) click of the mouse. Upon viewing the corresponding MCL, the student made his/her selection by another single (left) click of the mouse, which automatically placed his/her selection in the
corresponding blank node or link. This interactive computer-based concept map (see Figures 6 and 7) is called the Multiple-Choice Concept Map (MCCM) as its design borrows from both concept map assessment and traditional multiple-choice assessment.

**Issue with the Placement of the Selection List**

The second potential area of improvement of traditional SAFI concept maps, pertaining to the spatial placement of the list(s) of concepts and/or relations, is made evident by research from cognitive science (Chandler and Sweller, 1991, 1992; Cerpa, Chandler, & Sweller, 1996). Traditional SAFI concept maps, whether used for instructional purposes or assessment purposes (formative or summative), typically present the list of concepts and relations from which students have to select-and-fill-in the partially-completed expert map, separate from the concept map itself, usually to the side or on top of it (see Figures 3, 4, and 5). If the traditional SAFI maps are on paper as opposed to on a computer, these lists are often on separate sheets of paper, and, in the case of poor design, even on the backside of the concept map sheet. The limitations of working memory mentioned earlier (Baddeley, 1986; Miyake & Shah, 1999) make it difficult for a student to cognitively process the information on the map and the information on the list simultaneously. Thus, the student is forced to expend valuable and limited cognitive resources by intentionally directing his/her attention back and forth between the two separate sources of information (the map and the list) in order to complete the assessment task of selecting items from the list and filling in the existing blank nodes or links in the map. This design limitation may unnecessarily increase the student’s total mental effort and overload his/her working memory or cognitive capacity.
The previously mentioned phenomenon has been extensively researched in the area of instructional design within the framework of Cognitive Load Theory (CLT). Chandler and Sweller (1991, 1992) and Cerpa, Chandler, and Sweller (1996) have shown that splitting one’s attention between two separate or non-integrated sources of information increases overall mental effort and, therefore, the potential for a cognitive overload; this phenomenon has been termed as the split-attention effect (Chandler & Sweller, 1991, 1992; Cerpa et al., 1996). No such parallel work yet exists in the area of educational assessment. Nonetheless, the real fact that human working memory or cognitive capacity is limited is a reality that transcends disciplines, domains, and level of expertise (or prior knowledge) and, according to Leighton and Gierl (2007), “[b]orrowing theories from cognitive psychology and importing them into educational measurement initiatives is possible” (p. 10). Thus, this study investigated this potential area of improvement of traditional SAFI concept maps by using instructional design research findings from cognitive science to guide exploratory assessment design.

**Improving the Placement of the Selection List**

To avoid the negative consequence of increased cognitive load or mental effort attributed to the split-attention effect, Sweller (1999) suggests that, if text is essential to intelligibility, placing it on the diagram rather than separating it, reduces the cognitive load associated with having to split the learner’s attention between the two. Based on Snow and Lohman’s proposition that “different slices across the field of cognitive psychology might be needed to inform test design and evaluation” (1989, p. 265), this study employed this instructional design recommendation to guide the development of SAFI concept map assessment design. This was done in an attempt to determine whether
the detrimental consequence of increased mental effort or cognitive load caused by the split-attention effect makes its appearance in the select-and-fill-in concept map assessment method.

Both the traditional list (TL = 10 to 20 options/alternatives) and the multiple-choice list (MCL = 4 options/alternatives) were presented both non-integrated or separate from or adjacent to the concept map (see Figures 5 and 6) and integrated or inside the concept map (Figures 7 and 8). Naturally, all selection lists were viewed only at the student’s request by a single (left) click of the mouse and selections were made by another single (left) click. A higher overall mental effort or cognitive load on the student due to the non-integration of the selection list (TL or MCL) into the traditional SAFI map or the MCCM respectively, would confirm the existence of the split-attention effect in either or both select-and-fill-in concept map assessments (traditional SAFI and MCCM). Similarly, a decrease in the examinee’s overall mental effort due to the integration of the selection list (TL or MCL) would confirm the legitimacy of the weakness of not integrating selection lists into the concept map, indicating that integrated concept map assessment is superior to non-integrated concept map assessment.

Significance of Study

This research study is significant because it is guided by well-documented research recommendations to bridge the gap between cognitive psychology and educational measurement in an attempt to advance educational assessment (Mislevy, 2008; Leighton & Gierl, 2007; Nichols, 1994; Messick, 1989; Snow & Lohman, 1989). This research study is also timely, as it investigated concept mapping assessment when
the National Assessment Governing Board (NAGB) of the U.S. Department of Education is recommending the use of such assessments nationwide in their report entitled *Science Framework for the 2009 National Assessment of Educational Progress* (NAEP, 2008, p. 102).

This study addressed several different issues pertaining to educational assessment. First, concept maps in educational assessment has been shown to reflect valid and reliable measurement (Keppens & Hay, 2008; Clariana, Koul, & Salehi, 2006; Yin & Shavelson, 2004, 2008; Klein, Chung, Osmundson, Herl, & O'Neil, 2002; Ruiz-Primo, Schultz, Li, & Shavelson, 2001; Schau et al., 2001; McClure, Sonak, & Suen, 1999) in assessing students’ knowledge structures or schema-based knowledge (Pellegrino et al., 2001). Thus, using this type of assessment may enable the examiner to make more accurate inferences about the student’s ability to make connections and understand the relationships among concepts, a mainstream skill among expert performers.

Second, improving traditional SAFI concept maps by addressing their design weaknesses as specified earlier may render the Multiple-Choice Concept Map (MCCM) capable of minimizing, if not eliminating, the unnecessary mental strain during assessment. According to Grimley and Banner (2008) and Pickering and Gathercole (2004), increased demands to working memory negatively affect test performance. Therefore, it is of great importance to make more of the examinee’s limited cognitive capacity readily available for essential assessment cognitive processes by eliminating unnecessary processing of construct/content-irrelevant information (Mislevy et al., 2010; Baker, 2007; Stout, 2002). MCCMs’ efficient use of students’ limited cognitive capacity would prove invaluable because it would decrease the chances of the student (expert or
novice) becoming frustrated and cognitively overwhelmed due to poor assessment design; thus, the end result would yield scores that more accurately measure the student’s knowledge. Properly designed concept map assessments may indeed prove to be external knowledge representations with much desired characteristics of being explicit, efficient and valid (Mislevy, Behrens, Bennett, DeMark, Frezzo, Levy, Robinson, Rutstein, Stanley, Winters, & Shute, 2007).

Third, Huff and Goodman (2007) highlight the importance of reducing the time lag between test taking and score reporting, especially in large-scale summative assessments. They argue that the usefulness of test results/scores decreases as the delay in score reporting increases. This negative correlation impacting the usefulness of test results can be avoided altogether by reporting the score and even providing corrective feedback immediately after the student’s completion of the assessment activity. The test score and corrective feedback for the computer-based interactive Multiple-Choice Concept Maps (MCCMs) can easily be provided to the students by displaying the expert map on the screen or making it available for printing or downloading upon completion of the assessment. More traditional exams, including large-scale assessments, do not usually even provide attained scores to the examinee immediately after completing the test.

Research Questions and Hypotheses

The primary objective of this study is to evaluate the effects of (1) the type of the list(s) of concepts and/or relations and (2) the spatial placement of these list(s) on the examinee’s overall mental effort or strain. The study participants were assigned to four different treatment groups (SAFI-non-integrated using TL; SAFI-integrated using TL;
MCCM-non-integrated using MCL; MCCM-integrated using MCL) and compared on length of time to complete the assessment and examinee’s rating of personal mental effort exerted during the assessment. The participants’ pre-test scores on an 8-question multiple-choice test were used to control for prior knowledge. A 2x2 Factorial MANCOVA was used not only to determine the effect of each factor (the type of list/map used and the spatial placement of the list) on the composite dependent variable or the linear combination of the two dependent variables, but also to evaluate the effect of each factor on each of the two dependent variables (time on task and mental effort rating). The analysis also tested for an interaction effect. A Simple Planned Comparison was conducted to evaluate estimated mean differences between the primary group of interest (MCCM-integrated Group 4) and each of the other three groups. Of additional interest were differences among the four different groups on the examinees’ attitudes and impressions toward their respective assessment method; these data provided more feedback on the appropriateness of design for each treatment group.

**Research Question 1 and Hypothesis 1**

What is the effect of the type of list/map used (TL/SAFI and MCL/MCCM) on the length of time to complete the assessment and on the examinees’ rated mental effort? It was hypothesized that students assessed with the MCCMs (using the MCL) would complete the assessment task faster and exert less mental effort during the process than students assessed with the tradition SAFI concept maps (using the TL), indicating a minimized overall cognitive strain or load attributed to the superiority of the MCL over the TL.
Research Question 2 and Hypothesis 2

What is the effect of the spatial placement of the selection list (integrated and non-integrated) on the length of time to complete the assessment and on the examinees’ rated mental effort? It was hypothesized that students assessed with the integrated concept maps would complete the assessment task faster and exert less mental effort during the process than students assessed with the non-integrated concept maps, indicating a minimized overall cognitive strain or load attributed to the superiority of integrating the selection list into the concept map over not integrating it into the concept map.

Research Question 3 and Hypothesis 3

Does Group 4 (MCCM-integrated using MCL inside the map) take less time to complete the task and exert less mental effort during the assessment than the three other groups? It was hypothesized that MCCM-integrated Group 4 would take less time to complete the task and exert less mental effort during the assessment than all three other groups.

Research Question 4 and Hypothesis 4

How do examinees’ attitudes and impressions toward their respective concept map assessment activity differ among the four different treatment groups? It was hypothesized that examinees using the integrated MCCM would be more satisfied with their respective treatment than the other three groups (SAFI-non-integrated, SAFI-integrated, and MCCM-non-integrated).
Definition of Terms

**Concept Map** (Figure 1 and 2) – According to O’Neil & Klein (1997), concept maps are graphical representations of students’ understanding of interrelated concepts. They contain labeled nodes (circles or bubbles), which are connected by labeled links (or labeled lines). The nodes are labeled by concepts and the connecting links/lines are labeled by the relationships among the concepts.

**Integrated** (Figures 7 and 8) – is used to describe the spatial placement of the concept map selection list (whether TL or MCL). The concept map (whether SAFI or MCCM) is considered integrated when the selection list (TL or MCL) is presented or displayed within the concept map, right over top of the corresponding blank node or link.

**Link** – a part of a concept map that is in the shape of a line (usually directional) that links or joins together two nodes. These lines are typically labeled by one or more words that describe(s) the relationship between the concepts it links or joins together. Multiple such links, lined labels, or labeled lines together with the linked counter parts (nodes/concepts) are make up what is called the concept map.

**MCCM = Multiple-Choice Concept Maps** (Figures 6 and 7) – is a computer-based interactive concept map that makes use of MCLs to assess students’ knowledge. The student/examinee interacts with this type of concept map by using the computer mouse to first view the desired node/link-specific MCL and then to make the desired selection. This type of concept map can be integrated (Figure 7) or non-integrated (Figure 6), depending on the spatial placement of the MCL.

**MCL = Multiple-Choice List** (Figures 6 and 7) – is a list of no more than 4 concepts or relations that is carefully tailored to one blank node or blank link. These four carefully-
chosen alternatives (whether concepts or relations) contain one single correct answer and three (or two or one where appropriate) incorrect-but-plausible alternative(s) to the corresponding blank node or blank link. This type of selection list is used by the interactive computer-based MCCM and the examinee can choose to view it through a single (left) click of the mouse button and, upon viewing it, the desired selection can be made by another single (left) click of the mouse button.

**Node** – a part of concept map in the shape of a geometric figure (bubble, circle, oval, square, triangle, etc.) that contains or is labeled by a concept. Multiple such nodes, most often in circular shapes, are joined or linked together by links/lines (most often directional) to make up what is called the concept map.

**Non-integrated** (Figures 5 and 6) – is used to describe the spatial placement of the concept map selection list (whether TL or MCL). The concept map (whether SAFI or MCCM) is considered non-integrated when the selection list (TL or MCL) is presented or displayed on the side or next to the concept map, but on the same computer screen.

**SAFI = Select-And-Fill-In** – is used to describe the type of concept map assessment activity where the student or examinee is asked to select a concept or link from a list of concepts and relations and fill that selection in a concept map shell that is usually only partially filled. This type of concept map assessment, whether administered by paper and pencil/pen or on a computer, it always uses the traditional list (TL).

**TL = Traditional List** (Figures 5 and 8) – is a list of approximately 10 to 20 concepts, relations between concepts, or a combination of both (concepts and relations). This type of list, typically used in traditional SAFI concept map assessment, is used to fill in multiple blank nodes and/or blank links into a partially completed concept map. When
the concept map assessment task is administered by paper and pen/pencil, this type of list is usually presented either on the same sheet of paper as the concept map (next to or on top), on a separate sheet of paper, or (in the case of poor design) on the back page of the concept map page. When administered on computer, as was the case in this study, it is typically presented on the computer display/screen next to the concept map.
CHAPTER 2

REVIEW OF LITERATURE

The purpose of this chapter is to review the literature in order to make explicit the need for this study and reveal the conceptual or theoretical framework for the conducted research investigation. To this end, this chapter has been split into five major sections. The first section (The Development of Educational Assessment in the USA) provides a background on the development of educational assessment in the USA in order to contextualize contemporary educational assessment and communicate an appreciation for its present state and future direction. The second section (Concept Maps as an Alternative Assessment Method) explores research findings pertinent to concept map assessment in an attempt to impart a thorough understanding for this form of assessment. A sub-section of this chapter section covers comparative research studies that make evident why concept maps became one of the favorite alternative assessment methods over other forms of assessments, and then subsequent sub-sections narrow in on the concept map assessment method of interest to this research – the select-and-fill-in (SAFI) concept map assessment method. The third section (The Need to Bridge the Gap Between Cognitive Science and Educational Assessment) documents research recommendations that reveal the primary reason for the interest to further investigate the potential improvement of traditional SAFI concept map assessment. The fourth section (Cognitive Implications for Traditional SAFI Concept Maps) focuses on relevant cognitive research findings that help detect two potential weaknesses or areas of improvement, and makes explicit the need for exploratory research to inform further development of traditional SAFI concept map assessments. The fifth section of this chapter (Cognitive Science and Psychometrics
Inform Educational Assessment Design) uses both cognitive psychology and psychometric research findings and recommendations to develop the conceptual or theoretical framework for improving traditional SAFI concept maps.

The Development of Educational Assessment in the USA

Historical Developmental Milestones Lead to Excessive Use of Standardized Testing and of Multiple-Choice Items

During the early beginnings of the United States of America, access to varied forms of education was limited mainly due to the great distances and poor communication between the different states, which did not allow for the establishment of a school system. Hence, educational assessment during this time was casual and personal, usually limited to the testing of one or, in rare cases, a few students by one traveling teacher. Following the establishment of the first state Board of Education in 1837, the configuration of the American educational institution was starting to take shape through the efforts of the board’s secretary Horace Mann. Horace was instrumental at attracting students to the educational classrooms of Massachusetts and at seeing to the education and professional development of teachers (Jones, Jones, & Hargrove, 2003).

The formation of schools with actual age-appropriate grade levels and assessment practices in place further contributed to the formalization of teaching strategies and assessment methods, thus strengthening the newly establish education system. Although educational assessment during that time was mainly in the form of progress reports that were often highly subjective, the first documented achievement tests were administered in the mid-1800s. These achievement tests, although initially intended for individual
evaluation, were later modified to serve a different purpose. The great influx of immigrants to the United States, following setbacks of the Civil War (1861-1865), called for achievement tests to become standardized tools with which to “measure whether all children were receiving an equitable education” (Jones et al., 2003, p. 14).

In the late 1880s, Joseph Mayer Rice abandoned his medical practice and, after studying psychology and pedagogy abroad, returned to the United States and contributed to educational assessment foundations by conducting the first major comparative studies of students’ academic achievement through use of standardized testing in spelling (1897 – 33,000 subjects), in arithmetic (1902 – 6,000 subjects), and in language (1903 – 8,000 subjects). While Rice earned himself an unpleasant reputation among many professional educators for having exposed to the public the shortcomings and inefficiencies of the American school system of those times, other educational leaders such as Edward L. Thorndike, shared many of Rice’s beliefs and vision for the future of instruction and assessment in the American education system (Altenbaugh, 1999).

Edward L. Thorndike was a true pioneer in educational research. Based on his experimental and statistical investigations, he published several books and other publications on education-related topics ranging from teaching, learning, and measurement practices to administrative and social implications for teaching and learning. While John Dewey and others influenced the philosophy of education, Thorndike and those whom he inspired wrote reading and arithmetic books for students, pedagogical books for teachers and tests. Although “[m]any of the standardized achievement tests developed in the early 1900s in the U.S. were created by Thorndike-
trained measurement specialists” (Popham, W. J., 2000, p. 17), other standardized scales were concomitantly being developed abroad.

In 1905, Alfred Binet, a French physiological psychologist, created the first successful standardized scale for measuring intelligence. The scale was composed of thirty items and was the product of more than fifteen years of careful investigations and experimental research on children. Subsequent revisions of the scale appeared in 1908 and 1911 and a number of the items in the original scales are still included in the latest (1960 and 1986) revisions of the test. His collaborator in the development of these scales was Simon Theodore, who had earned a medical degree and was an intern at an institution for mentally challenged children when he associated himself with Binet in the 1890s. Further revisions of these scales made them available for wide use not only in schools, but also in industries and the army.

The widely acclaimed success of Binet’s effective intelligence tests stimulated much interest among U.S. psychologists who, then, contributed to the evolution of these tests by eliminating potential subjectivity/bias present at scoring, and by making them administrable to large groups of people, as opposed to only individuals. These two features of the standardized intelligence test (objective scoring and efficient administration) were accomplished at the request of the U.S. military who, faced with World War I, decided to use “scientific ways to maximize the efficiency with which it used human capital for its war machine” (Kennedy, 2003, p. 2).

The American Psychological Association (APA), then headed by Robert Yerkes, took on the challenge of “developing an objective and scientific way for planners to allocate men to positions in the military hierarchy” (Kennedy, 2003, p. 2). The APA
accomplished this by creating the Army Alpha test (for those who could read) and Army Beta test (for non-readers) that were designed to measure the mental ages of war recruits and volunteers. The results of these tests that were administered to over 1.7 million young men (most of whom were assessed with one of the five different forms of the Army Alpha) were used to categorize the examinees for various posts. Their unprecedented success was largely attributed to the tests’ intrinsic qualities of (1) effective and efficient administration to large groups of examinees and (2) objective and efficient scoring.

The above-mentioned qualities of the tests were mainly attributed to the then-new type of questions the tests contained – the multiple-choice-single-selection style question, presently known as the multiple-choice item. Following the success of its first large-scale implementation, the written form multiple-choice item used in the Army Alpha test became a major contributor to the development of the American educational assessment system and to the growth and development of standardized testing. The explosive growth in the number of published group aptitude and achievement tests following World War I drew some criticism and controversy during the 1920s. Nonetheless, the use of standardized testing in education continued to increase, especially after World War II, when public education in America grew considerably as evidenced by the radical increase in high school student enrollment and curriculum expectations (Kennedy, 2003).

During the 1960s, often defined as an “era of social, cultural, and racial disruption” (Jones et al., 2003, p. 14), the Elementary and Secondary Education Act of 1965 was devised in an attempt to minimize the differences in educational outcomes between students of different socioeconomic backgrounds. The stipulation that schools demonstrate effectiveness by use of standardized tests further propelled the increased use
of standardized testing. By this time, the use of multiple-choice standardized testing was also starting to be stimulated by technological developments such as the computer. Criticism of educational standardized testing during those times revolved mainly around its relevance for classroom decisions. In addition, concerns were expressed about the facts that “most standardized tests only partially matched local curriculum and provided little information about the skills and abilities students actually had” (Jones et al., 2003, p. 4).

As a result of previously-mentioned and other general concerns about the quality of education and educational assessment, standardized testing started to undergo certain changes; focus was placed on assessing what students could and could not do and on certifying whether students met minimum performance standards. These changes toward criterion-referenced and minimum competency testing transformed the nature of standardized tests and promoted mandatory large-scale testing, which translated into yet another increase in the use of standardized tests.

The 1983 publication by the National Commission on Excellence in Education (NCEE), *A Nation at Risk: The Imperative for Educational Reform*, pointed to low performing schools as a threat to national security in the face of a perceived comparative decline in the American economy. The widely held belief that the low performing American schools were somehow responsible for the perceived economic trouble fostered an unprecedented trust in the validity of test scores as measures of quality. Student results on the mandated high-stakes tests were linked not only to student promotions, but also to teacher and school evaluations, thus, often leading to questionable practices. For example, since teacher and school evaluations hinged on student performance on the
state-mandated high-stakes tests, teacher and often school administrators were motivated to do whatever was required to make sure that the students perform well on these tests; this, at times, meant that teachers provided inappropriate preparatory help to students and/or administrators reported inaccurate student results. (Kennedy, 2003, p. 5)

Criticism around test use and test preparation drew attention to the effect of testing on the teaching and learning process. By the mid-1990s, these discussions took the form of the authentic testing movement. The authentic testing movement changed the nature of standardized tests yet again. The traditional multiple-choice item, which had been the hallmark of standardized tests since the Army Alpha and Beta tests, was being “replaced by open-ended items, performances and other tasks thought to be more consistent with the complexity of the ways in which children think and learn” (Kennedy, 2003, p. 6). The standards movement holds that tests should not only “be more consistent with the ways in which people learn, but the content of the test and the criterion for performance should both reflect the highest standards with respect to national and international goals and norms” (p. 6).

Of noteworthy importance to the present state of the standardized testing is the No Child Left Behind (NCLB) Act of 2002, which is a reauthorization of the aforementioned Elementary and Secondary Education Act of 1965. Standardized testing, which is at the heart of the No Child Left Behind legislation, is used by the federal government as a tool to ascertain the good use or the poor use of federal funds by the school districts. Never before have student outcomes on standardized tests been so explicitly linked to significant consequences for schools and educators. Consequences of “recognition and rewards to schools that meet growth targets and interventions and the prospect of closure
for those who do not” (Kennedy, 2003, p. 7) were implemented in an attempt to meet the primary goal of managing federal funds more effectively and efficiently. Controversy of the NCLB legislation is primarily focused on the unintended consequences of the bill, such as lack of funds to suffering school districts.

As suggested throughout this sub-section of this chapter section, the extensive, and sometimes exclusive, use of multiple-choice items is what made the enormous and rapid growth of large-scale standardized testing in United States even possible, as multiple-choice items endowed standardized tests with the desirable characteristics of (1) effective and efficient administration and (2) objective and efficient scoring. However, research in the measurement and cognitive sciences during the past few decades and especially during the past 25 years, unveiled serious concerns about multiple-choice type of assessment. Thus, the next sub-section of this chapter outlines both psychometric and cognitive concerns about the still-widely-used multiple-choice method of assessment in an attempt to make explicit the rationale for the most recent changes in educational assessment, namely the move toward alternative assessment methods and practices.

The Rationale behind Alternative Assessments

The purpose of this sub-section of the chapter is to list and explain the implications of noteworthy psychometric and cognitive research findings that make evident the limitations and shortfalls of the widely-used multiple-choice type assessment. This, in turn, unveils the rationale behind the move toward alternative assessment methods and practices. As could be seen from the previous sub-section of this chapter, historical milestones such as the Elementary and Secondary Education Act of 1965, the 1983 publication A Nation at Risk: The Imperative for Educational Reform, and even the
more recent No Child Left Behind Act of 2002, all contributed to the increased administration of large-scale standardized tests while sharing the same underlying fundamental purpose – to use test scores to make inferences about students’ knowledge, skills, or abilities and to guide decision making about future instruction and even about students, teachers, and schools (Kennedy, 2003; DePascale, 2003). Consequently, increased stakes attached to test scores led to increased scrutiny toward the usability of test scores.

The need for meaningful test results, which would enable stakeholders to make valid inferences about areas of student misconceptions or about quality of teaching, initiated psychometric developments of sophisticated item analysis techniques and measurement models, such as the Item Response Theory or IRT (Waller, 1989). Through use of IRT, the combination of item clustering by skill or knowledge area and use of statistics generates valuable information about the levels and patterns of knowledge of students by tapping into latent variables behind the actual test items. This model, however, is too sophisticated for those not versed in statistics to employ and it provides results that are difficult to interpret. Thus, the usability of test scores of multiple-choice assessment is greatly diminished by the absence of elaborate statistical analyses skills.

Further addressing the usability of test scores, in *Test Theory Reconceived* Mislevy (1996) also made a strong case about multiple-choice tests’ inability to provide useful information to educators for enhancing teaching. He pointed out that simple test statistics, such as mean, median, and mode, provide educators no information about which area of teaching needs improvement. Basic item analysis mainly focuses on calculating and comparing percentages of students who answered particular items.
correctly and percentages of students that chose each of the wrong answers or distracters, while considering which students (high-scorers or low-scorers) answered each item correctly. When more low-scorers answer an item correctly, the item is considered to be bad and is usually tossed out. Correspondingly, when more high-scorers answer an item correctly, the item is considered to be a good discriminant, and is usually kept. This type of item analysis only provides information about the quality of the test and the items themselves with little or no valid inference about area of student misconceptions or quality of the teaching. Thus, Mislevy (1996) introduced the idea of assessing students’ knowledge structures by mapping a network of probabilities of skills-possession and corresponding test items that would provide the instructor with visual feedback about areas where students lack skills to complete tasks. This proposed visual networked feedback about students’ knowledge, however, requires much time and mathematical/statistical expertise of classroom instructors.

Advances in the science of cognition also raised concerns about the types of skills multiple-choice assessment can measure. In Test Theory Reconceived (1996), Mislevy expressed his apprehension for traditional multiple-choice tests’ inability to measure more than students’ fragmented understanding. He identified that traditional multiple-choice tests are based on differential and behaviorist theories and that they could not assess students’ organization of knowledge or knowledge structures, something cognitive psychology has identified to be a key element to knowledge acquisition and use. He encouraged incorporating cognitive psychology into educational measurement, as did Messick (1989), Snow and Lohman (1989), and Nichols (1994).
The educational assessment community understood the need to align educational assessment with contemporary cognitive theories and, thus, when the focus of school reform driven by promoting higher order thinking skills shifted to the state level by the early 1990s, most of the nearly 85 percent of the states that still chose multiple-choice tests for their statewide assessments did in fact combine these with other forms of performance assessments (Jones et al., 2003). According to Jones and colleagues (2003) “[d]uring the 1993-1994 school year, thirty-eight states assessed writing, twenty-five states used other performance assessment, and seven states required portfolios. Two states, Kentucky and Maine, had abandoned multiple-choice testing altogether in favor of alternative assessment strategies” (p. 16).

The continued inclusion of alternate forms of assessments were later addressed on a more formal and large-scale basis by The Committee on the Foundations of Assessment, established by the National Research Council (NRC) in 1998 “to review and synthesize advances in the cognitive sciences and measurement and to explore their implications for improving educational assessment” (Pellegrino et al., 2001). The work of this committee, often described to be a natural extension of How People Learn (the 1999 report by the Committee on Developments in the Science of Learning), was funded by grants from the National Science Foundation (NSF). A 3-year long study conducted by 18 experts with diverse perspectives on educational assessment generated the comprehensive report of 2001 called Knowing What Students Know. Based on much investigation of top-notch developments in the fields of psychometrics, cognitive psychology, and educational technology, this report clearly listed and thoroughly
described educational assessment strengths and weaknesses and made recommendations on how to improve educational assessment.

The Committee on the Foundations of Assessment (Pellegrino et al., 2001) confirmed earlier concerns of Messick (1989), Snow and Lohman (1989), Nichols (1994), and Mislevy (1996) by reiterating the limitation and sometimes inability of multiple-choice educational assessments (1) to capture the complex knowledge and skills emphasized in contemporary standards (e.g., students’ organization of knowledge) and (2) to provide useful information to educators to help enhance teaching. According to the Committee on the Foundations of Assessment, the enduring and still wide use of multiple-choice assessment denotes that educational assessment is still “derived from early theories that characterize learning as a step-by-step accumulation of facts, procedures, definitions, and other discrete bits of knowledge and skills” (Pellegrino et al., 2001, p. 60). Furthermore, although tests that utilize well-constructed multiple-choice items can be reliable and measure complex cognitive processes (p. 194), “when assessment fails to provide information that can enhance learning, it leaves educators ill-equipped to close achievement gaps” (p. 29).

Thus, the confirmed deficiency of multiple-choice assessment encouraged the continued increase in the use of alternative forms of performance assessments. However, many alternative assessment types still lack the ability to capture complex knowledge and skills (Pellegrino et al., 2001), such as “coherent structures of knowledge” (p. 23) or “organization of knowledge” (p. 27), and to provide useful feedback to the evaluator. In addition, many alternative assessments are also time and cost inefficient when compared
to multiple-choice assessment. The next sub-section of this chapter outlines some of the more prominent types of alternative assessment.

Alternative Assessments

Among the numerous alternative assessment types that are thought to require “the active construction of meaning rather than the passive regurgitation of isolated facts” (McMillan, 2004, p. 15) are: essay writing, open-ended questions, interviews, think-alouds, concept maps, performance assessments, portfolios, and journals. Many believe that these assessment types are “consistent with cognitive theories of learning and motivation as well as societal needs to prepare students for an increasingly complex workplace” (p. 15). However, most alternative assessments are time and cost inefficient and, therefore, impractical for use both at the classroom level and in large-scale assessments. For example, even though interviews and think-alouds may, in fact, inform the evaluator about examinees’ complex knowledge structures and cognitive processes, the time and cost inefficiency for administering and scoring these two forms of assessment renders them unfeasible for use even at the classroom level. Therefore, interviews and think-alouds have been used primarily for research purposes as they can give deep insight into an examinee’s mind. Even essays and open-ended questions, have proved to be quite challenging to implement, as teachers are faced with time-consuming administration and the challenges of designing and employing sophisticated objective rubrics. For example, administering and scoring essay exams takes considerably longer than administering and scoring multiple-choice tests, most often scored by electronic scantron scanning equipment. According to Banks (2005), the time-intensive administration and scoring of many alternative assessments leave teachers little and
maybe not enough time to teach knowledge relevant to traditional standardized tests their students must take.

The time and cost inefficiency of many alternative assessments became even more evident when alternative assessment types were included in large-scale standardized tests. According to O’Neil and Klein (1997), it costs about $5.00 to score a student written essay using the Iowa Tests of Basic Skills (ITBS). While this compares well to the estimated $90.00 per student/test to score a hands-on performance measure, it is still an expensive alternative to multiple-choice testing. Another challenge of essay questions is that the accuracy of the assessment of an examinees’ understanding of a concept may be affected by the examinees’ writing ability level (Banks, 2005). Nonetheless, the undeniable need to assess students’ schematic knowledge structures (Pellegrino et al., 2001) sustained the continued increase in the use of essays and open-ended questions even in large-scale assessments, in spite of their time and cost inefficiency. For example, in 2005 the Scholastic Assessment Test (SAT) included an essay question as part of its college aptitude test (Banks, 2005).

This steady increase in the use of essays and open-ended questions in large-scale assessments is both a result of and a contributor to recent technological advancements that improve the time and cost efficiency of these types of assessments by computerized automatic scoring of text pieces from essay and open-ended questions. According to Pellegrino and colleagues (2001), the most widely used technological scoring systems are based on a cognitive theory of semantics called Latent Semantic Analysis (LSA). LSA involves “constructing a multidimensional semantic space that expresses the meaning of words on the basis of their co-occurrences in large amounts of text. Employing
mathematical techniques, LSA can be used to ‘locate’ units of text within this space and assign values in reference to other texts.” (Pellegrino et al., 2001, p. 269). Thus, this assessment system takes a student’s essay and compares it either to a set of pre-graded essays at varying quality levels or to one or more model, expert-written essay(s). It can provide not only a single overall score, but multiple scores based on which essay it is compared to.

Evaluation studies for these LSA based computerized essay scoring tools point to reliability coefficients as high as those for scores assigned by pairs of human raters (Pellegrino et al., 2001, p. 269). Banks (2005) points out that the evaluators of the Graduate Management Admissions Test (GMAT) found that the human raters of the Analytical Writing Assessment essay question for the GMAT and the corresponding e-rater, which is a “computer program that scores essays based on syntax, organization, analysis and usage of standard English”, generally agreed 87 to 94 percent of the time (p. 153). However, despite such positive evaluations of electronic or computer-based essay scoring systems, an inescapable reality about essay and open-ended assessments is that detection of schematic knowledge structures and cognitive processes is highly dependent on the examinee’s language skills.

**Concept Maps as an Alternative Assessment Method**

Concept maps, the so-called “tools for organizing and representing knowledge” (Novak, 2003, p. 1), have been used for educational assessment purposes almost as long as they have been used for teaching and learning. Developmental and validation research findings over the past few decades have encouraged their use in learning and assessment.
This section of the chapter covers research findings relevant to the development of concept map assessment methods and their establishment among the favored alternative assessments (NAEP, 2008).

**Defining Concept Maps**

“It has been well acknowledged that [assessing schematic knowledge structures] should be the cornerstone of the assessment design process” (Kalyuga, 2005). The undisputable need to assess students’ “coherent structures of knowledge” (Pellegrino et al., p. 23) or “organization of knowledge” (p. 27) is what prompted the creation of so many alternative assessments and sustained their increased use even in spite of previously mentioned time and cost inefficiencies. Concept maps are among the favored alternative assessments, because they make examinees’ knowledge visible without burden to limited working memory capacity, which according to Mislevy and colleagues (2007), is characteristic of effective external knowledge representations.

According to O’Neil and Klein (1997), concept maps are graphical representations of students’ understanding of interrelated concepts. They contain labeled nodes (circles or bubbles) that are connected by labeled links (labeled lines or lined labels). The nodes are labeled by concepts and the connecting links/lines are labeled by the relationships among the concepts. Concept map assessment methods or techniques differ significantly based on the method of administration (paper-and-pencil or computer-based) and the type of assessment task (construct-from-scratch or select-and-fill-in). For example, a student might be given paper and a pencil and be instructed to construct a concept map from scratch on the topic of introductory statistics. The student then would bring together all that he/she knows about the mentioned topic and organize it on paper.
by placing all known concepts inside nodes/bubbles and labeling the joining lines with known relationships between those concepts. Other teachers or evaluators may choose to provide the student with a computer-based partially filled concept map and a list of concepts and/or relations and instruct the student to fill in the blank nodes and/or links by using the selection list provided. Regardless of the assessment task, however, concept maps visually evaluate students’ knowledge organization or schema-based knowledge structures in a given domain without being hindered by language dependence, as are essays and other alternative assessments. While concept maps are not as language-dependent as other forms of assessment, concept mapping also faced initial time and cost inefficiency challenges.

**Overcoming Initial Time and Cost Inefficiencies**

The primary factor causing the time and cost inefficiency of concept maps was the accurate scoring of paper-and-pencil concept maps. The existence of multiple concept map assessment methods or techniques lead to multiple concept map scoring methods. Among the different concept map scoring methods suggestions are the following: Novak and Gowin (1984) proposed counting of hierarchies or cross-links among them, Walker and King (2003) recommended calculating node-to-link ratios, McClure and colleagues (1999) suggested counting number of correct node-link relationships and assigning scores based on levels of correctness, while O’Neil and Klein (1997) and O’Neil, Chung, and Herl (1999) proposed evaluating nodes, links and node-link relationships by comparing them to those on an expert map.

A study by McClure and colleagues (1999) compared six different concept map scoring methods and correlated students’ map scores to similarity scores with a master
map created by a teacher. In their study, relational scoring in conjunction with a master map yielded the highest reliability coefficient (.76) and the highest concurrent validity (.61). In this method, raters assigned scores 1 through 3 to each proposition based on correctness level as compared to a master map constructed by a teacher. The researchers concluded that the preferred scoring method for student-constructed paper-pencil concept maps as assessment tools should be relational scoring in conjunction with a master map.

Using technology to score concept maps increases time and cost efficiency (O’Neil & Klein, 1997; O’Neil et al., 1999; Clariana et al., 2006; Scalise & Gifford, 2006), as has been the case for multiple-choice scoring and other types of assessment. When O’Neil & Klein (1997) explored the feasibility of using machine-scored paper-and-pencil-administered concept maps for large-scale assessments versus performance assessments or multiple-choice item based tests, they did a cost-benefit analysis which indicated that scoring time and associated costs are very high for scoring performance assessment and low for scoring essays and even lower for scoring concept maps.

O’Neil and Klein (1997) believed that even though it would take only minutes to familiarize a student with a computer-based concept map assessment task, it would still be too expensive and, in some cases impossible, to test all students using computer-based concept map assessments; this is why they developed the CRESST software that can score the scanned pre-printed pencil-and-paper-teacher-administered concept maps. It might be noteworthy to mention here that during the present time, assessing all students through the use of computers is not only possible, but is often the norm. Despite the obligatory shipping costs and software charges, the estimated cost of fifty cents ($0.50) to score one pre-printed pencil-and-paper concept map compares very well to the estimated
ninety dollars ($90.00) to score a performance assessment and the estimated five dollars ($5.00) to score an essay test.

Klein and colleagues (2002) developed an algorithms-based software that scores students’ concept maps by comparing them to characteristics found in two model concept maps constructed by experts. This software awards students a half a point if a link explanation matches one of the expert maps and a full point if it matches with both expert maps. Klein and colleagues added another rating method, which awarded three points for propositions that experts identified as ‘critical,’ and one point for any valid proposition that experts did not include in their maps. Inter-expert agreements of propositions included in the maps were .72 and .77 for two different topics investigated (hearing and vision respectively). The two concept map scoring methods yielded correlations of .92 and .84, which caused the researchers to conclude that the original simpler method would suffice.

Clariana and colleagues (2006) confirm the effectiveness and efficiency of using automatically derived concept map scores for educational purposes. Their study required graduate student participants to read about and research online the structure and function of the heart and circulatory system and then construct a concept map and write a 250-word essay on the topic. In this study, the human-rater concept map scores correlated moderately high (r = 0.75) only on term agreement with an expert. All of the computer-derived scores, however, were significantly correlated to human-rater essay scores and to the earlier-mentioned LSA computer-based essay scores by a maximum correlation coefficient (r) of 0.83 (p. 323). These results indicate that the a cost efficient “self-scoring
concept map tool should be of great interest to the research and education communities” (p. 323).

Validity and Reliability

The two psychometric properties of validity and reliability are essential to the establishment of a solid assessment method (Ruiz-Primo & Shavelson, 1996); thus, the validity and reliability of concept maps as assessment methods have been explored extensively. For example, the usability of scores obtained from various concept map assessment methods has been analyzed by comparing examinee maps with teacher-constructed master maps (McClure et al., 1999; Kinchin, Hay, & Adams, 2000; Kankkunen, 2001; Walker & King, 2003). Other concept map assessment validation research investigated the relationship between scores achieved on concept map assessments and scores achieved on other types of assessments, such as multiple-choice or essay and open-ended question tests (O’Neil & Klein, 1997; Rice, Ryan, and Samson, 1998; Cawley, Zimmaro, van Meter, & Theodorou, 1999; Schau et al., 2001; Ruiz-Primo et al., 2001; Tsai, Lin & Yuan, 2001; Yin, Vanides, Ruiz-Primo, Ayala, & Shavelson, 2005; Yin & Shavelson, 2004, 2008). However, in order to fully appreciate the details of research findings by the above-mentioned researchers, this sub-section of the chapter defines the types of validities and reliabilities used to validate various concept map assessment methods.

Validity, as defined by Messick (1989), is “an integrative evaluative judgment of the degree to which empirical evidence and theoretical rationales support the adequacy and appropriateness of inferences and actions based on test scores or other modes of assessment” (p. 5). Similarly, McMillan (2004) defines validity as “a characteristic that
refers to the appropriateness of the inferences, uses, soundness, trustworthiness, or legitimacy of the claims or inferences that are made on the basis of obtained scores” (p. 59). This means that, if inferences are to be made about students’ knowledge organization or schematic knowledge structures, concept map assessment methods need to generate scores that would render accurate inferences about the schematic knowledge structures the students possess.

Concurrent validity is a type of criterion-related validity, where “the meaning of the scores … [is] substantiated externally by appraising the degree to which empirical relationships with other measures, or lack thereof, is consistent with that meaning” (Messick, 1995). This implies that concept map assessment methods need to correlate highly with other more established and effective assessment types so as to ascertain the meaningfulness of obtained scores. This type of validity is essential to the establishment of any new form of assessment and many research findings have confirmed high levels of concurrent validity of concept map assessment methods with other traditional and non-traditional assessment types.

Reliability, as defined by McMillan (2004), is “concerned with the consistency, stability, and dependability of the scores. In other words, a reliable result is one that shows similar performance at different times” (p. 65). Thus, if concept maps are to become sound assessment methods, they need to be reliable, which means that they need to produce dependable scores that are stable and replicable. Since the purpose of this chapter is not to go into in-depth explanations of all the different sources of reliability evidence, the next paragraph introduces a few relevant reliability measures most often
calculated when investigating and comparing the reliability of different concept map assessment methods to other types of assessment methods.

A stability estimate of reliability, also called test-retest reliability, requires that the same assessment be administered to the same group on two different occasions separated by a specified period of time; correlating these two sets of scores gives a high or low reliability coefficient. “Internal consistency evidence is based on the homogeneity of the scores on items measuring the same trait” (McMillan, 2004, p. 69) and is relatively easy to obtain as it does not necessitate the development and/or administration of a second assessment/test. It must be noted, however, that the most common types of internal consistency estimates (split-half, Kuder-Richardson’s KR20 and KR21, and Coefficient Alpha) were originally designed for use with tests and not with concept maps. Another common reliability measure is inter-rater reliability, also called scorer or rater consistency, where the scores of one rater are correlated with the scores of another rater in an attempt to identify measurement bias due to rater characteristics, such as fatigue, halo effect, and others.

Comparing Concept Map Assessment to Other Types of Assessments

Ground-breaking research by Rice and colleagues (1998) provided evidence that concept maps, as compared to matched multiple-choice tests using the same assessment blueprint, can provide valid assessment of knowledge and comprehension level learning outcomes, as well as higher level cognitive structures. In their study, seventh-grade students drew paper-and-pencil concept maps from scratch by using a list of concepts matched to the constructs and answer choices on a multiple-choice test. These concept maps were then graded based on the number of correct propositions. Inter-rater
agreement for the map scores was 98 percent and multiple-choice instrument reliabilities (KR-20s) on several of the tests ranged between .63 and .88. Correlations of .41 to .70 between concept map scores and multiple-choice test scores, as well as correlations between concept mapping scores and science, reading, and mathematics standardized test scores of .82, .84, and .82, respectively, indicate good concurrent validity evidence for concept maps as educational assessment methods.

Other researchers who looked at correlations between student-constructed concept map scores and multiple-choice test scores reported correlation coefficients that ranged from -.30 (Cawley et al., 1999) to .70 (Rice et al., 1998; Tsai et al., 2001). It is important to note, however, that even though scoring methods for the student-generated concept maps varied among the researchers, Schau and colleagues (2001) point out that the researchers who found low or negative correlations (e.g.: Cawley et al., 1999) did not specify the reliabilities of the multiple-choice exam scores or of the concept map scores, which varied substantially.

Kinchin and colleagues (2000), however, claim that student-constructed maps, when qualitatively analyzed, can show student knowledge in context, rather than isolated, as it would be captured by more traditional assessments such as multiple-choice tests. In addition, they claim that student-constructed concept maps can identify what students know and how the knowledge is represented in their minds and, thus, replace even time- and labor-intensive interviews. They suggest, however, that student-constructed concept maps should be qualitatively evaluated, with the goal of pinpointing misconceptions and diagnosing learning-related problems, rather than providing students with scores. They argue that the scoring of only valid links constructed by students disregards the value of
other links that are not quantitatively valid, but are necessary for students’ understanding. They believe that student-constructed concept maps show the perception of the map’s author, which reflects personal beliefs, biases, and topic knowledge - rather than memorized facts.

In 1997, O’Neil and Klein had found correlations of .70 between concept mapping scores and essay scores and suggested that concept maps may even have the potential to replace essay items in large-scale assessment tests. Klein and colleagues (2002) investigated the validity and reliability of concept map assessment by using a multi-trait multi-method validity matrix to look at correlations between three assessment methods (essays, multiple-choice tests, and computer-based concept maps) on two different contents (hearing and vision). They found the reliability level of .95 for the two essay raters on both topics. The inter-item alpha reliability levels were .77 for the multiple-choice questions on hearing and only .34 for the multiple-choice questions on vision due to the level of item difficulty. The multi-trait multi-method validity matrix indicated high to moderate correlations between essay and multiple-choice tasks (.71 and .53), moderate correlations between concept mapping and multiple-choice tasks (.46 and .33), and moderate correlations between concept mapping and essay tasks (.56 and .43).

Schau and colleagues (2001) found strong evidence of internal consistency and concurrent validity on select-and-fill-in the nodes (SAFI) concept maps. SAFI maps require students to fill in missing nodes on a teacher-constructed map. Middle school science and undergraduate astronomy students’ SAFI concept map scores were compared to their scores on multiple-choice exams. In addition, undergraduate students’ SAFI scores were also compared to their scores on the relatedness ratings (RR) technique,
which “elicits a student’s structural knowledge of a domain” (p. 138). SAFI scores increased from the beginning to the end of the semester, and correlated well with both the multiple-choice and the construct-from-scratch concept map scores. Internal consistency was also strong. More specifically, the Cronbach’s alphas were .92 and .91 for scores on the SAFI map measure and .81 and .79 for scores on the multiple-choice test for seventh and eighth graders, respectively. Cronbach’s alpha for the SAFI map scores was .83 for the undergraduate group. Correlations between students’ scores on the paper-and-pencil SAFI maps based on the percentage of correct propositions and percent correct on the multiple choice tests were .74 and .77 for students in the two grade levels. This indicates good concurrent validity levels for the SAFI concept map assessment method when compared with reliable multiple-choice item based assessments.

The correlations calculated by Schau and colleagues (2001) were likely higher than those found by other researchers because the select-and-fill-in (SAFI) format is more similar to multiple-choice assessment format, in that students have a predetermined or assigned list of concepts/relations from which they have to select-and-fill-in the desired answer into the appropriate blank on the partially-filled expert concept map. This is considerably different than having to rely on a list of concepts and a list of links/relations to build a map from scratch. It also differs substantially from concept map assessment methods where students have to fill in missing nodes and links from recall (Tsai et al., 2001). Schau and colleagues admit that a potential limitation of the SAFI method is that it provides the expert’s knowledge structure for the examinee, rather than allowing the examinee’s natural structures to surface as does the construct-from-scratch concept map method. On the other hand, the examinee can benefit from comparing his/her preexisting
knowledge structures to those of an expert and such a comparison may result in actual clarifications of preexisting subtle or fundamental misconceptions. In addition, the fact that the SAFI concept map assessment method used by Schau and colleagues is more consistent with the traditional method of multiple-choice testing is not a weakness, because it still incorporates the added benefit characterized by the use of concept maps – assessing students’ much-desired knowledge organization or schema-based knowledge structures.

Comparing Different Concept Map Assessment Methods or Tasks

Studies investigating the validity and reliability of different concept map assessment methods, tasks, or techniques have been conducted in the framework of generalizability theory, where multiple sources of error that can affect the psychometric properties of a single measure are considered (Shavelson & Webb, 1991). In concept map assessment research, these sources can include mapping techniques, raters, terms depicting concepts, and propositions (Yin et al., 2005). Most researchers who conducted studies investigating the dependability of concept maps in the generalizability theory framework did so through comparisons of different concept-mapping tasks, techniques, or methods (Yin et al., 2005; Yin & Shavelson, 2004, 2008).

Ruiz-Primo (2004) encapsulated existing concept mapping assessment techniques by placing them on a continuum from high degrees of directedness to low degrees of directedness. She describes the lowest degree of directedness as the activity of constructing a map from scratch while given no concepts, no linking labels, and no structure or shell. According to Chang, Sung, and Chen (2001), the construct-from-scratch concept mapping activity includes the least scaffolding since the student receives
very little direction and no directional pointers or help. The highest level of directedness, as depicted by Ruiz-Primo (2004), is described by the concept mapping task of either filling in the nodes or filling in the links into a partially-filled expert map from an assigned list of concepts or relations. The partially-filled expert map or shell that represents the organizational structure of an expert’s knowledge and the predetermined lists of concepts or relations serve as the props that provide scaffolding directing the user toward an already functional structure, organization, and representation of an expert’s knowledge (Chang et al., 2001). Ruiz-Primo indicates that techniques at the opposite ends of the continuum can even measure different types of examinee knowledge and surely serve different purposes.

Research findings by Kinchin and colleagues (2000) suggest that qualitatively comparing a student’s concept map and a subsequent teacher-student-collaborative concept map may be the most effective way of showing student's zone of proximal development. For their study, science students had to construct concept maps on the same content at different times throughout the semester. These multiple student-constructed concept maps were interpreted as having demonstrated progressive levels of student understanding, thus, suggesting the usefulness of these results to guide instruction.

Other researchers have also found qualitative evidence of the validity of concept mapping to demonstrate students’ construction of initial meanings and progression of meaning-making (Kankkunen, 2001; Walker & King, 2003). In their study, Walker and King (2003) found that faculty generated dense networks of concepts and applications while students generated fewer connections among concepts pertaining largely to domain content (p. 5). In addition, students’ maps varied qualitatively over time, where “later
maps used more precise vocabulary, were more coherently constructed, and contained more connections among concepts” (p. 8). They conclude that their findings suggest that “concept maps are a useful means of portraying the process of knowledge transformation from novice to expert” (p. 14).

In their study, Yin and colleagues (2005) compared student performance on two concept map assessment tasks, the construct-a-map with student-created linking phrases and the construct-a-map with selected or assigned linking phrases. Ninety-two eighth-graders in six middle-school science classes taught by the same teacher participated in the study on the content of density, mass, and matter. In the student-created linking phrase condition, students constructed maps relying on a list of concepts, but had to create their own linking phrases to place between the concepts. The latter concept map assessment technique provided students with both linking phrases and concept terms and students had to then assemble a concept map using the assigned concepts and links. In agreement with other findings (Ruiz-Primo et al., 2001), Yin and colleagues (2005) contend that the student-created linking technique more accurately reflects differences in students’ knowledge structures and preexisting misconceptions. This is so because students are not locked into a pre-constructed expert map like they are when they use a fill-in-the-node or fill-in-the-link technique, such as the one investigated by Schau and colleagues (2001).

Select-and-fill-in concept map assessment methods, on the other hand, provide better test-retest reliability levels than student-constructed concept maps, even when the concept maps include fewer propositions (Yin & Shavelson, 2004, 2008). The low test-retest reliability of student-constructed concept maps is partially due to some measurement bias present when scoring the student-constructed concept maps. The
selected linking (selected-and-fill-in) concept mapping technique, however, allows for objective computerized scoring of concept maps (Klein et al., 2002; Clariana, 2006), which renders high test-retest reliability (Yin & Shavelson, 2004, 2008). Hence, Yin and Shavelson suggest that the selected linking technique, or the so-called select-and-fill-in (SAFI) concept map assessment method, might be more appropriate for large-scale summative assessment, while the student-constructed linking technique (or the student-constructed concept map assessment method) is more appropriate for formative classroom assessment. This seems to be a logical assertion given that student-constructed concept maps have proven to be more sensitive to detect existing misconceptions (Ruiz-Primo et al., 2001; Yin et al., 2005) while select-and-fill-in (SAFI) concept maps better evaluate how the students’ knowledge compares to that of an expert which looks at a more comprehensive snapshot picture of the person’s knowledge. The fact that inferences made from SAFI concept map scores exhibit psychometric properties of high validity and reliability (Rice et al., 1998; Schau et al., 2001; Klein et al., 2002; Yin & Shavelson, 2005) strengthens the supposition that they have the potential to become effective and efficient summative assessment tools to be used in the classrooms (NAEP, 2008).

Schau and colleagues (2001), who found strong evidence of internal consistency and concurrent validity on traditional select-and-fill-in (SAFI) concept maps, list three limitations to student-constructed concept maps that SAFI maps overcome. (1) Examinees using student-constructed concept maps must learn how to draw concept maps and may perceive this activity to be time-consuming, tedious and frustrating. SAFI concept maps do not require this task. (2) Even though language dependence for concept map assessments in general is low when compared to essay or multiple-choice
assessments (Stoddart, 2006), student-constructed concept maps are higher on language
dependence when compared to SAFI maps. Assessments that are high in language
dependence, such as essays, place their validity at risk (Bailey, 2000) as they may
function as language proficiency tests rather than assessments of content knowledge
(Abedi, Leon, & Mirocha, 2003). The Standards for Educational and Psychological
Testing (American Educational Research Association, American Psychological
Association, & National Council on Measurement in Education, 1999), state that
assessments that are high on language dependence may introduce construct-irrelevant
components into the testing process and, thus, hinder their ability to accurately reflect the
competencies which the assessment intended to measure. Boscardin, Jones, Nishimura,
Madsen and Park (2008) confirm these potential negative effects of high language
dependence on testing outcomes in their report Assessment of Content Understanding
Through Science Explanation Tasks. (3) There is no universally accepted and simple
scoring system for student-constructed concept maps. In other words, traditional SAFI
concept maps eliminate the potential measurement error attributed to assessment task,
language dependence, or scoring method/system.

Think-aloud protocols and expressed opinions about the SAFI concept map
assessment method by students and teachers alike indicated that this method was easy to
use and that students filling out the SAFI maps used “strategies requiring connected
understanding” (Schau et al., 2001, p. 148). The SAFI concept map assessment method
also addresses well recommendations of the Committee of the Foundations of
Assessment to make “students’ thinking visible to both their teachers and themselves”
(Pellegrino et al., 2001, p. 4) and to provide useful feedback about and facilitate the measurement of students’ “coherent structures of knowledge” (p. 23).

In conclusion to this section on concept maps as an alternative assessment method, it is important to reiterate that traditional SAFI concept map assessment has gained recognition among alternative assessments because it meets contemporary educational assessment objectives (e.g., evaluating the organization of knowledge) and possesses desired psychometric properties (e.g., validity, reliability, etc.). Research findings from cognitive psychology (Grimley & Banner, 2008), however, indicate that the assessment design of traditional SAFI concept maps may impose unnecessary strain on the examinee’s limited cognitive capacity (Baddeley, 1986; Miyake & Shah, 1999). In order to fully appreciate the underlying cognitive implications in the process of assessment design and evaluation, however, it is important to first understand the current state of educational measurement or assessment and its relationship to cognitive psychology.

The Need to Bridge the Gap Between Cognitive Science and Educational Assessment

Mislevy (2008) describes the present state or relationship between cognitive science and educational assessment as a “widening chasm” that needs to be bridged (p. 124). Messick (1989) and Snow and Lohman (1989) had already showcased the need to bridge this gap between cognitive psychology and educational assessment a couple of decades ago in Robert Linn’s Educational Measurement (1989) in their respective chapters entitled “Validity” and “Implications of Cognitive Psychology for Educational
Measurement.” The ideas expressed in these two chapters influenced both novice and established scholars to further investigate the use of cognitive psychology to inform test development. This endeavor was additionally supported by sudden increases in high-stakes large-scale testing and accepted notions that assessment “innovations must be justified by providing something beyond what is available through standard formats” (Scalise & Gifford, 2006, p. 38).

Early initiatives to bridge the gap between cognitive science and educational assessment (Nichols, 1994) started to shape the theoretical framework for Cognitive Diagnostic Assessment (CDA), which holds that educational assessments need to be placed under the “cognitive microscope” (Leighton & Gierl, 2007, p. 10). CDA further purports that “the onus is on educational researchers to adapt our methods, techniques, and tools to incorporate cognitive theory and, more important, to actively create, modify, and test theories of cognition for educational measurement purposes” (p. 10). In other words, the 1989 proposition that “different slices across the field of cognitive psychology might be needed to inform test design and evaluation” (Snow & Lohman, p. 265) is the basis for many research initiatives in educational assessment design and evaluation.

Almost 20 years later, the National Assessment Governing Board (NAGB) of the U.S. Department of Education reiterated this need to bridge the gap between cognitive psychology and educational measurement in the report *Science Framework for the 2009 National Assessment of Educational Progress* (NAEP, 2008). One of the recommendations of the NAEP report (2008) in the chapter titled *Overview of the Assessment Design* is to use concept maps in classroom assessment. This recommendation is based on the stated opinion that “concept-mapping tasks tap a science
ability that is difficult to measure by other means” (p. 102). The next section of this chapter covers relevant research findings from cognitive science that reveal potential weaknesses of traditional SAFI concept maps and supposedly guide their continued improvement.

Cognitive Implications for Traditional SAFI Concept Maps

One of the recommended ways to bridge the gap between cognitive psychology and educational assessment is accomplished by using relevant research findings from cognitive science to guide test evaluation and assessment design (Snow & Lohman, 1989; Mislevy, 2008; NAEP, 2008). Putting traditional SAFI concept maps under the “cognitive microscope” (Leighton & Gierl, 2007, p. 10) in this way reveals two potential assessment design shortcomings that might impose unnecessary mental strain on the examinee’s limited cognitive capacity.

The Traditional Concept Map Selection List

The first assessment design weakness and potential area of improvement for traditional SAFI concept maps is directly related to the type of list(s) of concepts and/or relations to be selected-and-filled-in the partially filled expert concept map. The traditional list (TL) of concepts and/or relations used by traditional SAFI concept maps usually contains about 10 to 20 concepts, relations, or a combination of concepts and relations. The assessment task of the student/examinee is to select a concept or relation from the list and place/write/type/click it into the appropriate blank node or blank link within the concept map; naturally, this task is replicated with every attempt to fill in a blank node or blank link. As specified earlier in this chapter, SAFI concept maps
eliminate examinee frustration caused by learning how to construct a map and engaging in the task of constructing-from-scratch. Unfortunately, cognitive psychology research findings shed light on the fact that the use of the traditional SAFI concept map list might render the select-and-fill-in task time-consuming, frustrating, and even cognitively overwhelming.

Research findings in cognitive science (Baddeley, 1986; Miyake & Shah, 1999) describe working memory, the main information processor or the locus of control for our cognitive processes, to be limited in both capacity and duration when handling information. For example, early research on memory suggested that we can remember a series of approximately seven random numbers (Miller, 1956) for a mere several seconds (Peterson & Peterson, 1959), unless the numbers are intentionally rehearsed. Thus, processing 10 to 20 pieces of information on a traditional list (TL) used by traditional SAFI concept maps might cognitively overload the student’s limited working memory capacity. Grimley and Banner (2008) documented the negative effects of low working memory on educational achievement on the General Certificate of School Education (GCSE) exam, namely that students with low working memory capacity are at risk of attaining low GCSE scores. If current SAFI concept map assessment design imposes additional cognitive strain on the map user, then possibly all students’ scores are negatively affected, those with high and low working memory. Further consideration for the limitations of working memory has led researchers to encourage the design of assessments that eliminate content/construct irrelevant components (Mayer, 2001; Stout, 2002; Baker, 2007; Mislevy, Bejar, Bennett, Haertel, & Winters, 2010).
Processing 10 to 20 pieces of information on a traditional list (TL) used by SAFI concept maps may cognitively overload the student’s limited working memory capacity. When the examinee reviews the selection list of 10 to 20 concepts/relations to fill in a given blank node or link, most of the alternatives/options on the list are not related to the corresponding blank node or link as they are the correct answer to another blank node or link. In addition, the students/examinees have to view and review the same list, or sometimes lists, with every attempt to fill in a blank node or link, thus being forced to process the same pieces of information repetitively; this redundant task inadvertently and unnecessarily solicits invaluable, limited cognitive resources.

According to Stout (2002), complex assessment items such as concept maps should prevent the examinee from “doing irrelevant mental processing” (p. 109). If cognitive psychology is to inform the assessment design of SAFI concept maps, the traditional selection list of concepts/relations (TL) needs to change so as to avoid soliciting the examinee’s cognitive resources with content or construct irrelevant material (Mayer, 2001; Stout, 2002; Baker, 2007; Mislevy et al., 2010). A later section of this chapter explains proposed ways to improve the quality of the traditional list (TL), but the following sub-section presents the second potential weakness of traditional SAFI concept maps, in an attempt to provide a fuller understanding of the problem.

The Spatial Placement of the Concept Map Selection List

The second assessment design weakness and potential area of improvement of traditional SAFI concept maps, namely the spatial placement of the list(s) of concepts and/or relations, is made evident by research from cognitive science as well. Traditional SAFI concept maps, whether used for instructional purposes or assessment purposes
(formative or summative), typically present the list of concepts and/or links, from which students have to select-and-fill-in the partially completed expert map, separate from the concept map itself, usually to the side or on top of it. If the traditional SAFI concept maps are on paper as opposed to on a computer, these lists are often on separate sheets of paper, and, in the case of poor design, even on the backside of the concept map sheet.

The limitations of working memory mentioned earlier (Baddeley, 1986; Miyake & Shah, 1999; Grimley & Banner, 2008) make it difficult for a student to cognitively process the information on the map and the information on the list simultaneously. Thus, the student needs to expend valuable and limited cognitive resources by intentionally directing his/her attention back and forth between the two separate sources of information (the map and the concept map selection list) in order to complete the assessment task of selecting items from the list and filling in the existing blank nodes or links in the map. This design limitation might unnecessarily increase the student’s total mental effort and overload his/her working memory or cognitive capacity.

Chandler and Sweller (1991, 1992) and Cerpa and colleagues (1996) researched the previously mentioned phenomenon within the frameworks of Cognitive Load Theory (CLT) and Mayer (2001) within the frameworks of Multimedia Learning. Research findings by cognitive load theorists have shown that splitting one’s attention between two separate or non-integrated sources of information increases overall mental effort and, therefore, the potential for a cognitive overload; this phenomenon has been termed as the split-attention effect (Chandler & Sweller, 1991, 1992; Cerpa et al., 1996). Similar research within the context of multimedia learning by Mayer (2001) and Moreno & Mayer (1999) revealed that
In two of two tests, learners performed better on retention tests when corresponding text and illustrations were placed near each other on the page (or when corresponding on-screen text and animation segments were placed near each other on the screen) than when they were placed far away from each other. In five of five tests, learners performed better on transfer tests when corresponding text and illustrations were placed near each other on the page (or when corresponding on-screen text and animation segments were placed near each other on the screen) than when they were placed far away from each other. (Mayer, 2001, p. 81)

The spatial contiguity principle described in the above mentioned research findings suggests that visual material is best held and processed in working memory when it is presented in close proximity on the same page or screen. No such parallel work yet exists in the area of educational assessment. Nonetheless, the real fact that human working memory or cognitive capacity is limited is a reality that transcends disciplines, domains, and level of expertise (or prior knowledge).

Leighton and Gierl (2007) stated that, although difficult, “[b]orrowing theories from cognitive psychology and importing them into educational measurement initiatives is possible” (p. 10). According to Grimley and Banner (2008) working memory limitations affect test performance much the same way – when cognitive resources are used for irrelevant tasks such as “visually searching the page or screen” (Mayer, 2001, p. 81), test performance is negatively affected. Thus, if cognitive psychology is to inform the assessment design of SAFI concept maps, the spatial placement of the selection list of concepts/relations needs to be further investigated so as to ascertain whether the split-attention effect of instructional design makes its appearance in traditional SAFI concept map assessment and, if so, determine the ideal placement of the selection list.

The previously mentioned weaknesses of traditional SAFI concept maps indicate that deeper investigation of these two potential areas of improvement is necessary in
order to further improve this type of assessment method. Thus, the next section of this chapter describes proposed changes to traditional SAFI concept maps and reveals the theoretical motivators for the conducted research.

Cognitive Science and Psychometrics Inform Educational Assessment Design

This section of the chapter describes this researcher’s attempt to “adapt [educational assessment] methods, techniques, and tools to incorporate cognitive theory and, more important, to actively create, modify, and test theories of cognition for educational measurement purposes” (Leighton & Gierl, 2007, p. 10). More specifically, cognitive theories from working memory (Baddeley, 1986; Miyake & Shah, 1999; Grimley & Banner, 2008), multimedia learning (Mayer, 2001) and cognitive load (Chandler and Sweller, 1991, 1992; Cerpa et al., 1996) are used to guide exploratory research in the area of assessment design for SAFI concept maps.

The Multiple-Choice List (MCL) Used by the Multiple-Choice Concept Map (MCCM)

As described earlier, the traditional list (TL) used by traditional SAFI concept map is a potential weakness because of the high number of options or alternatives (10 to 20) it contains. Cognitively processing all 10 to 20 options simultaneously and repetitively with every attempt to fill in a blank node or link increases the examinee’s overall mental effort. This increased mental effort can be avoided by eliminating the necessity to process non-related alternatives. Therefore, this study used relevant research findings from cognitive psychology and psychometrics to adjust the assessment design of
tradition SAFI concept maps so as to eliminate the unnecessary portion of the mental effort attributed to the use of the traditional list (TL).

Research findings from cognitive science dealing with working memory (Miller, 1956; Peterson & Peterson, 1959; Baddeley, 1986; Grimley & Banner, 2008), human cognitive architecture (Sweller et al., 1998; Kalyuga et al., 2003) and multimedia learning (Mayer, 2001) suggest that the number of options (whether concepts or relations) contained in a list be reduced by excluding all non-related alternatives, so as to avoid unnecessary cognitive processing. Psychometric research findings on the optimum number of alternatives/options for multiple-choice questions (Waller, 1989; Hutchinson, 1997; Abad et al., 2001) suggest that the lists include no more than four concepts or relations to choose from. Thus, this study investigated the use of the multiple-choice list (MCL), so named by this researcher because of its similarities to the list of multiple-choice alternatives used by multiple-choice items.

The aforementioned carefully-chosen alternatives (whether concepts or relations) contained one single correct answer and three incorrect-but-plausible alternative(s) to the corresponding blank node or blank link. Therefore, each blank node and each blank link had its corresponding carefully-tailored multiple-choice list (MCL) of four alternatives. Inadvertently, to present these numerous multiple-choice lists (MCLs) nearby or adjacent to the concept map simultaneously, whether on paper or on a computer screen, would be ineffectual as they might visually and cognitively overwhelm the student. However, in computer-based concept map assessment as was the case in this study, this potential visual clutter was easily avoided through the use of background programming that enabled the student to view only the multiple-choice list (MCL) that corresponded to the
blank node or blank link that the student chose by a single (left) click of the mouse. Upon viewing the corresponding MCL, the student made his/her selection by another single (left) click of the mouse, which, in turn, automatically placed his/her selection in the corresponding blank node or link. This interactive computer-based concept map was named by this researcher the Multiple-Choice Concept Map (MCCM) because of its resemblance to both the concept map selection list (MCL) and the traditional multiple-choice item alternatives list. A decrease in the examinee’s overall mental effort due to the use of the multiple-choice list (MCL) would confirm the legitimacy of the weakness of the traditional list (TL), indicating that the multiple-choice concept map (MCCM) is superior to the traditional SAFI concept map.

**The Integration of the Selection List (TL and MCL) into the Concept Map**

As described earlier in this chapter, spatially placing/presenting the concept map selection list separate from or non-integrated into the actual concept map poses a potential weakness because it might forces the examinee to split his/her attention between the information on the concept map and that on the selection list repetitively with every attempt to fill in a blank node or link, thus causing an increase in overall mental effort. This increased mental effort could be avoided by eliminating the necessity to split the examinee’s attention between two separate or non-integrated sources of information (the map and the list). Therefore, this study used relevant research findings from cognitive psychology on human cognitive architecture to adjust the assessment design of select-and-fill-in concept maps so as to eliminate the unnecessary portion of the mental effort attributed to the split-attention effect, so termed by extensive research in the area of instructional design.
To avoid the negative consequence of increased cognitive load or overall mental effort attributed to the split-attention effect, Sweller (1999) suggests that, if text is essential to intelligibility, placing it on the diagram rather than separating it, would reduce the cognitive load associated with having to split the learner’s attention between the two. Similarly, to avoid the negative effect of the spatial contiguity principle on overall mental effort, Mayer (2001) suggests that pieces of information that need to be processed together need to be presented on the same screen, close together, and simultaneously. Based on Snow and Lohman’s proposition that “different slices across the field of cognitive psychology might be needed to inform test design and evaluation” (1989, p. 265), this study employed these instructional design recommendations to guide the development of SAFI concept map assessment design in an attempt to determine whether the detrimental consequence of increased overall mental effort or cognitive load caused by the split-attention effect makes its appearance in the select-and-fill-in concept map assessment method as well.

Thus, both the traditional list (TL = 10 to 20 options/alternatives) and the multiple-choice list (MCL = 4 to 2 options/alternatives) have been presented both integrated (within the concept map, right over top of the blank node or link) and non-integrated (next to but separate from the concept map, on the right side). Naturally, the multiple-choice lists (MCLs) were viewed only at the student’s request by a single (left) click of the mouse and selections were made by another single (left) click. A higher overall mental effort on the examinee due to the non-integration of the TL or the MCL into the traditional SAFI map and the MCCM respectively, would confirm the existence of the split-attention effect in either or both select-and-fill-in concept map assessments.
(traditional SAFI and MCCM). A lower overall mental effort due to the integration of the selection list (TL or MCL) would confirm the legitimacy of the weakness of not integrating selection lists into the concept map, indicating that integrated concept map assessment is superior to non-integrated concept map assessment.

In conclusion, this chapter has covered the development of educational assessment and made evident the rationale for its continuous evolution and its present state of multiple alternative assessments. Then, it covered in-depth concept maps as an alternative assessment method and narrowed in on the SAFI concept map assessment method. The section on the need to bridge the gap between cognitive science and educational assessment reveals the primary reason for the interest to further investigate potential improvement of SAFI concept map assessment. The section on the cognitive implications for traditional SAFI concept maps focused on cognitive research findings that helped detect two potential weaknesses or areas of improvement and made explicit the need for exploratory research to inform further development of SAFI concept map assessments. The last section of this chapter used both cognitive psychology and psychometric research findings and recommendations to develop the conceptual or theoretical framework for improving SAFI concept maps. The next chapter explains the methodology this research study used to test the legitimacy of the previously described weaknesses of traditional SAFI concept maps and the effectiveness of the integrated and non-integrated multiple-choice concept maps (MCCMs).
CHAPTER 3

METHOD

The purpose of this study was to examine two potential areas of improvement in the assessment design of traditional SAFI concept maps by evaluating the effects of (1) the type of selection lists of concepts and relations and (2) the spatial placement of these lists on the examinee’s overall mental effort. The methodology for how that was accomplished is described in this order: Participants and Setting, Procedure, Materials, Instruments and Measures, Research Design, Research Questions. The hypothesized results are listed in the last section of this chapter.

Participants and Setting

The G*Power software was used to determine an appropriate sample size for the proposed research (Faul, Erdfelder, Lang, & Buchner, 2007). G*Power is an open source statistical software primarily used for power analyses. An A Priori Power Analysis calculation given an error probability value ($\alpha = 0.05$), power ($P = 0.95$), and effect size ($f^2 = 0.12$) revealed the need for a total sample size of 80 participants ($N = 80$). Thus, this study used 87 undergraduate students from a public and a private mid-western college. The participants of this investigation completed this study as an extra credit activity for their respective statistics course. The participants had full control over whether or not they chose to participate in this particular study and over whether or not they wanted to release their data for research purposes.
Procedure

Following the approval by the IRB boards from each university and college (see Appendix 1), professors at the two colleges were contacted by the researcher via email and phone to offer the study as a potential extra credit activity for their students. The professors/instructors who agreed to offer this study as an extra credit activity for their students introduced the study and told the students the time and place (computer lab) where the study would be conducted. To ensure students’ high interest levels to participate in this research study, the description of the study found in the Participant Recruitment Form informed the students that this study would involve interactive concept map assessments on a topic that is relevant to their course content material and that an expert concept map on the content used in the research would be given to the all participants following the completion of data collection for the study. Students were informed that this map can serve as corrective feedback to their existing knowledge organization and be used for studying during the duration of their course and even as a reference resource throughout their degree.

When the participants reported at the computer lab, the researcher guided each participant to sit at one of the computers with a stable broadband internet connection running a Javascript-enabled browser at a minimum recommended screen resolution of 1024x768 pixels or higher. Once all the participants were seated, the researcher introduced himself and the study and reminded everyone that participation was voluntary and that extra credit would be given them according to their respective professor’s specifications. The participants were then given the Informed Consent to read and sign, after which 3-digit Participant IDs were randomly assigned to all the participants through
the sequential distribution of small pieces of pre-cut paper that contained their respective 3-digit ID and the direct URL to the study’s website. The 3-digit Participant ID dictated which of the different treatment groups they were randomly assigned to. Once the participants logged into the study’s website using their respective 3-digit ID, they first completed the Demographics Questionnaire (see Appendix 2) and the 8-question multiple-choice pre-test (see Appendix 3) to measure their prior knowledge on the content matter. A very simple (non-elaborate) practice interactive, web-based concept map assessment activity on the topic of cats and dogs was administered for a timed duration of 2 minutes to familiarize the participants with their respective assessment method (see Figures 9, 10, 11, and 12). After this practice task, participants were automatically taken to their respective concept map assessment activity. Upon completion of their respective concept map assessment task, participants rated their mental effort attributed to the assessment design using the online Mental Effort Rating Scale (see Figure 13) and completed the online Attitudes and Impressions Questionnaire (see Appendix 4) about their respective concept map assessment method.

Materials

Two independent variables (type of list and spatial placement) were used to create the four different treatment groups. Although participants were randomly assigned to each of the four different groups, a pre-test (see Appendix 3) was administered to all participants to ascertain their levels of prior knowledge on relevant content matter. The pre-test initially contained 9 multiple-choice questions on the topic of correlation. However, one of the question stems mistakenly included an error, which rendered the
question alternatives incorrect and confusing; this was confirmed by the participant responses and so the question containing the error was excluded and the participants’ scores adjusted accordingly. These 8 items have been used extensively by an instructor who taught this content matter and item analyses have shown the pre-test items to be reliable. Cronbach’s Alpha for the 8 items for the 87 participants was .75 ($M = 4.62$, $SD = 2.33$), confirming a good level of reliability. The 8-question pre-test designed for this research tested the same concepts and misconceptions as the concept map assessments that followed.

A different concept map assessment method based on the same assessment blueprint on the topic of correlation was administered to each group. The master/expert concept map constructed for this study that served as the assessment blueprint for all four concept map assessment methods can be seen in Figure 1. Group 1 received the traditional SAFI concept map (Figure 5) that used the traditional list (TL) displayed on the right side of the concept map (non-integrated). Group 2 received a quasi-traditional SAFI concept map (Figure 8) that used the TL, but displayed within the concept map (integrated). Group 3 received the MCCM (Figure 6) that used the multiple-choice list (MCL) displayed on the right side of the concept map (non-integrated). Group 4 received the MCCM (Figure 7) that used the MCL, but displayed within the concept map (integrated).

All concept maps were interactive and designed to fit in a monitor display with minimum screen resolution of 1024x768 pixels, as most monitors and computers comfortably support this screen resolution. Participant responses were compiled and recorded to the server through the inconspicuous use of web-based programming.
languages. Javascript was used to compile the data in hidden fields and Perl was used to write the data to the server in a format that is easily importable by Excel and PASW Statistics (formerly SPSS) software for statistical analysis.

All the concept maps were identical in content in order to ensure uniform content validity (see Figures 5, 6, 7, and 8). They all contained the same blank nodes and links in their initial presentation. The assessment task was essentially identical in that the examinees were required to use their respective list (TL non-integrated, TL integrated, MCL non-integrated, MCL integrated) to fill in each blank node and link. This was accomplished by two single (left) clicks of the mouse – one click on the desired blank node or link activated the selection list and one click on the desired option/alternative automatically filled in the blank node or link. Examinees were able to change their selections until they chose to submit their final and fully completed version of the concept map.

Traditional SAFI concept maps, similar to the ones used by Schau and colleagues (2001), have thus far been administered primarily by paper and pencil (see Figure 3). Similarly to Chang and colleagues (2001), the traditional SAFI-non-integrated maps and the accompanying TL used by this research study were administered on computer (see Figures 4 and 5) in an attempt to ensure that the assessment task remained the same for all 4 groups. As described in Chapter 2, the MCL used by the MCCM-integrated and MCCM-non-integrated groups (see Figures 6 and 7) differed from the TL in that the MCL contained only 4 options/alternatives, which were carefully-tailored to each respective blank node or link. The TL comprised 13 options/alternatives to contain the correct answer to all the blank nodes or links. When non-integrated, both lists (TL and
MCL) were displayed on the right side of the map (Figures 5 and 6). When integrated, both lists (TL and MCL) were displayed right over top of the corresponding blank node or link within the concept map (Figures 7 and 8). As described in Chapter 1 and 2, the MCCM differed from the more traditional SAFI concept map in that the MCCM (whether non-integrated or integrated) used the MCL instead of the TL.

**Instruments and Measures**

Since the different treatments of this research study were compared on their effects on the participants’ mental effort or cognitive load, it is essential to acknowledge the different types of cognitive load. The Cognitive Load Theory (Chandler & Sweller, 1991, 1992; Sweller & Chandler, 1991; Sweller, 1994; Sweller, van Merrienboer, & Paas, 1998; van Gerven et al., 2000; Kalyuga et al., 2001) as it relates to instructional design defines cognitive load as the total amount of mental activity that takes place in working memory when engaging in any mental task or activity. The four general principles about the cognitive load theory, as described by Cooper (1998) are that: (1) working memory is limited, (2) long term memory is virtually unlimited, (3) anything that involves mental processing of information must happen in working memory before it can be encoded or re-encoded into long term memory, and (4) exceeding the working memory resources will render learning (and by extension any mental task) ineffective. Cognitive load is of three different types: Intrinsic, Extraneous, and Germane.

**Types of Cognitive Load**

Intrinsic cognitive load is determined by the mental demands of the task at hand and is increased or reduced by the variation of interacting elements or pieces of
information that need to be processed together for effective learning. Intrinsic cognitive load cannot be controlled by either the learner or the instructor (Chandler & Sweller, 1991, 1992; Sweller & Chandler, 1991).

Extraneous cognitive load is imposed on the learner by the design of the materials or instruments. Extraneous cognitive load can be controlled by the designer of the materials. Ideally, good design contains few or no extraneous factors or distracters in an attempt to minimize extraneous cognitive load and free up working memory resources (Chandler & Sweller, 1991, 1992; Sweller & Chandler, 1991).

Germane cognitive load is the processing capacity used to construct schemata and store it in long-term memory, which is what establishes any learning actually occurred. Unlike intrinsic and extraneous cognitive load, which are imposed on the learner, germane cognitive load is usually voluntary and, in the case of highly motivated learners, very controlled (Sweller et al., 1998).

Measuring Mental Effort or Cognitive Load

Research on human cognitive architecture within the framework of instructional design theory (Paas & van Merrienboer, 1993; Paas, van Merrienboer, and Adam, 1994) indicates that mental effort can be measured by: (1) learner performance (on primary and secondary tasks), (2) length of time spent on task, (3) objective physiological measures such as changes in pupil diameter, heart rate, body temperature, and breathing rate, and (4) subjective measures in the form questionnaires usually administered to the learner after the task is completed (Paas, Tuovinen, Tabbers, & van Gerven, 2003). To evaluate the effects of the previously mentioned independent variables (the type list/map and the spatial placement of the list) on the examinees’ overall mental effort, this research study
compared the four groups on the length of time to complete the assessment and on the
rated mental effort.

The Mental Effort Rating Scale

According to Kalyuga, Chandler, and Sweller (2001), subjective rating scales are
the preferred instruments for measuring mental effort because they correlate highly with
objective rating measures (.80 to .99) and they are non-intrusive and do not disturb task
performance as do some objective physiological measures. This study used, as one of its
mental effort measures, the non-intrusive subjective 9-point Likert-type scale that
contains one item ranging from 1 (very very low) to 9 (very very high). This popular
unidimensional one-item subjective rating scale was found to be more useful than an
objective heart rate measure (Paas & van Merrienboer, 1994) and sensitive to relatively
small differences in mental effort (Paas et al., 2003). When utilized to ascertain
differences in rated mental effort, this instrument was “valid, reliable, and unintrusive”
(Paas et al., p. 66). Hypothesizing that similar working memory challenges might occur
during educational assessment as a cause of the design of the assessment, the Likert-type
mental effort rating scale in this study used the following wording:

Please take the time to THOROUGHLY CONSIDER the overall mental effort
you exerted or the mental strain you experienced while completing the interactive
concept map on the topic of ‘Correlation’. Using the scale below, please rate (AS
ACCURATELY AS POSSIBLE) ONLY your mental effort caused by the MAP
DESIGN user-friendliness or flaws (Figure 13).

The participants were asked to rate only the mental effort caused by the map
design (see the scale’s wording) because the purpose of this study was to examine
potential assessment design weaknesses. According to Chandler and Sweller (1991,
1992), weaknesses in the design of instructional materials and instruments are the cause
of extraneous cognitive load. Similarly, weaknesses in the assessment design of traditional SAFI concept maps may have an effect on extraneous cognitive load, and some of this study’s treatment conditions may reflected improved assessment design.

**Attitudes and Impressions Questionnaire**

To investigate differences in examinees’ attitudes and impressions toward their respective concept map assessment method, a questionnaire was administered to all participants immediately following the completion of the assessment activity. This questionnaire collected quantitative data through use of multiple-choice type questions as well as qualitative data thorough use of open-ended questions.

**Research Design**

The four different groups (SAFI-non-integrated using TL; SAFI-integrated using TL; MCCM-non-integrated using MCL; MCCM-integrated using MCL) were compared on length of time to complete the assessment and examinees’ rating of personal mental effort exerted during the assessment. The participants’ pre-test scores were used to control for potential differences in overall mental effort due to prior knowledge of the content matter (topic of statistical correlation). A 2x2 Factorial MANCOVA was used to determine the effect of each factor (the type of list/map used and the spatial placement of the list) on the composite dependent variable or the linear combination of time on task and mental effort rating. The Tests of Between-Subjects Effects of the Multivariate General Linear Model ran using PASW Statistics 18.0 software provided valuable information relative to the effect of each factor or independent variable on each of the two dependent variables. Though no hypothesis was made about an interaction effect, the
analysis also tested for an interaction. To evaluate mean differences between the primary group of interest (MCCM-integrated Group 4) and each of the other three groups (see Research Questions 3), a Simple Planned Comparison was conducted. Differences among the four different groups on the examinees’ attitudes toward their respective assessment method were evaluated using participants’ responses to the Attitudes and Impressions Questionnaire containing both quantitative and qualitative questions.

Research Questions

The previously-mentioned research design provided data to answer the following research questions: (1) What is the effect of the type of list/map used (TL/SAFI and MCL/MCCM) on the length of time to complete the assessment and on the examinees’ exerted mental effort? (2) What is the effect of the spatial placement of the selection list (integrated and non-integrated) on the length of time to complete the assessment and on the examinees’ exerted mental effort? (3) Does Group 4 (MCCM-integrated using MCL inside the map) take less time to complete the task and exert less mental effort during the assessment than the other three groups? (4) How do examinees’ attitudes and impressions toward their respective concept map assessment activity differ among the four different treatment groups?

Hypothesized Results

(1) It was hypothesized that students assessed with the MCCMs (using the MCL) would complete the assessment task faster and exert less mental effort while doing it than the students assessed with the traditional SAFI concept maps (using the TL), indicating a
minimized overall cognitive strain or load attributed to the superiority of MCL over TL. (2) It was hypothesized that students assessed with the integrated concept maps would complete the assessment task faster and exert less mental effort during the assessment than students assessed with the non-integrated concept maps, indicating a minimized overall cognitive strain or load attributed to the superiority of integrating the selection list into the concept map over not integrating it into the concept map. (3) It was hypothesized that MCCM-integrated Group 4 would take less time to complete the task and exert less mental effort during the assessment than all three other groups. (4) It was hypothesized that examinees would favor the integrated MCCM over the other three (non-integrated MCCM, integrated SAFI, and non-integrated SAFI).
CHAPTER 4

RESULTS

The data for this study were gathered using both quantitative and qualitative measures. The Pre-test, the Interactive Concept Maps, and the Mental Effort Rating Scale were all quantitative and the Attitudes and Impressions Questionnaire contained both multiple-choice and open-ended questions. This chapter first presents the demographic information collected from the participants and then organizes the quantitative results based on the research questions of this study. The last section of this chapter is dedicated to the analysis of the qualitative data collected through the Attitudes and Impressions Questionnaire.

Demographic Information

The analyses were based on 87 participants from two mid-western colleges, more specifically, 64 (73.60%) from a community college, and 22 (25.30%) from a private college; one person did not provide college information. Their average age was 22.62 (SD = 5.60). Among the 87 participants were 60 females (69.80%), 26 males (30.20%), and 1 participant who did not report gender. The majority of participants identified themselves as Caucasian (73 or 83.90%), with grade levels between sophomore and senior (75 or 87.20%). The participants were evenly distributed through the four treatments, with 20, 22, 22, and 23, as the total number of participants in each different group (see Table 1).
Research Questions

The research design and data collection and analysis aimed to provide answers to the following research questions: (1) What is the effect of the type of list/map used (TL/SAFI and MCL/MCCM) on the length of time to complete the assessment and on the examinees’ exerted mental effort? (2) What is the effect of the spatial placement of the selection list (integrated and non-integrated) on the length of time to complete the assessment and on the examinees’ exerted mental effort? (3) Does Group 4 (MCCM-integrated using MCL inside the map) take less time to complete the task and exert less mental effort during the assessment than the other three groups? (4) How do examinees’ attitudes and impressions toward their respective concept map assessment activity differ among the four different treatment groups?

Quantitative Results

A 2x2 factorial MANCOVA was administered with dependent variables Time on Task and Mental Effort Ratings, and factors Type of List and Spatial Placement. The Mean for Time on Task was 271.78 seconds (SD = 84.09) and the Mean for Mental Effort Ratings scores was 4.46 (SD = 2.07). All the means and other descriptive statistics can be seen in Table 2. As can be seen in Table 3, the Pearson correlation between the prior knowledge covariate and time on task was -0.21 (p < .05) and that between prior knowledge and mental effort was -0.37 (p < .01). The two dependent variables, time on task and mental effort, had a moderately strong positive correlation of 0.53 (p < .01), which is desirable for a Factorial MANCOVA. The prior knowledge covariate pre-test score on the 8-question multiple-choice pre-test on the topic of correlation was
statistically significant, $F(2, 81) = 6.95, p = .00, \eta^2 = .15$, observed power = .92 (see Table 4). The pre-test was reliable with a Cronbach’s Alpha level of .75. Though no hypothesis had been made about an interaction effect, the analysis tested for it and found no statistical significance, $F(2, 81) = 0.86, p = .43, \eta^2 = .02$, observed power = .19 (see Table 4).

(1) What is the effect of the type of list/map used (TL/SAFI and MCL/MCCM) on the length of time to complete the assessment and on the examinees’ exerted mental effort?

The 2x2 Factorial MANCOVA results (see Table 4) indicated that there was a statistically significant main effect of the type of list, $F(2, 81) = 3.20, p = .05, \eta^2 = .07$, observed power = .60, on the linear combination of time on task and mental effort ratings. Thus, 7% of the variance in the dependent variables can be explained by type of list, and that one can expect the same result 60% of the times for similar studies. More specifically, however, the test of between-subjects effects (see Table 5) showed that there was a statistically significant difference, $F(1, 82) = 6.46, p = .01, \eta^2 = .07$, observed power = .71, between SAFI and MCCM on mean time on task, but not on mean mental effort ratings, $F(1, 82) = 1.42, p = .24, \eta^2 = .02$, observed power = .22. This means that 7% of variance in time spent on the assessment can be explained by the type of list/map used, and one can expect this same result 71% of the times when similar studies are conducted. As can be seen in Table 2, those with MC concept maps (estimated $M = 250.76, SE = 11.98$) completed the assessment 43.87 seconds faster than those with SAFI maps (estimated $M = 294.63, SE = 12.41$), but the mental effort ratings gathered by the 9-point Likert-type scale differed only by 0.50 points. The estimated mean for the MC
groups was 4.22 ($SE = 0.29$) and the estimated mean for the SAFI groups was 4.72 ($SE = 0.30$). Standard deviations and true means for these groups on time on task and mental effort are also provided in Table 2. Though no hypothesis had been made about an interaction effect, the analysis tested for it and found no statistical significance, $F(1, 82) = 0.70, p = .41, \eta^2 = .01$, observed power = .13 (see Table 5).

(2) What is the effect of the spatial placement of the selection list (integrated and non-integrated) on the length of time to complete the assessment and on the examinees’ exerted mental effort?

The 2x2 Factorial MANCOVA results (see Table 4) showed that there was no statistically significant main effect of the placement of the list, $F(2, 81) = .61, p = .54, \eta^2 = .01$, observed power = .15) on the linear combination of time on task and mental effort ratings. More specifically, the test of between-subjects effects (see Table 5) showed that there was no statistically significant difference between the non-integrated groups and the integrated groups on mean time on task, $F(1, 82) = .66, p = .42, \eta^2 = .01$, observed power = .13 and on mean mental effort ratings, $F(1, 82) = .07, p = .79, \eta^2 = .00$, observed power = .06. Those with integrated lists (estimated $M = 265.71, SE = 11.98$) completed the assessment only 13.97 seconds faster than those with non-integrated (external) lists (estimated $M = 279.69, SE = 12.42$), and the mental effort ratings gathered by the 9-point Likert-type scale were only 0.11 apart. The estimated mean for the integrated groups was 4.52 ($SE = 0.29$) and the estimated mean for the non-integrated groups was 4.41 ($SE = 0.30$). Standard deviations and true means of these groups on time on task and mental effort are provided in Table 2. Though no hypothesis had been made about an interaction effect, the analysis tested for it and found no statistical significance, $F(1, 82) = 0.70, p = .41, \eta^2 = .01$, observed power = .13 (see Table 5).
effect, the analysis tested for it and found no statistical significance, $F(1, 82) = 0.23, p = .64, \eta^2 = .00$, observed power = .08 (see Table 5).

(3) Does Group 4 (MCCM-integrated using MCL inside the map) take less time to complete the task and exert less mental effort while completing it?

To evaluate mean differences between the primary group of interest (MCCM-integrated Group 4) and each of the other three groups on both dependent variables (time on task and mental effort), a Simple Planned Comparison was conducted to determine whether or not the MCCM integrated Group 4 spent less time on task and exerted less mental effort than the other three groups. Mean Differences were obtained by subtracting the Group 4 estimated means on time on task and on mental effort from the estimated means of each of the other 3 groups. The mean differences on time on task between Group 4 MCCM-integrated and the other three groups were 57.85 seconds ($SE = 24.56$) for Group 1 SAFI-non-integrated, 58.32 seconds ($SE = 23.97$) for Group 2 SAFI-integrated, and 28.42 seconds ($SE = 24.05$) for MCCM-integrated Group 3. The Simple Planned Comparison results (see Table 6) indicated that there was a statistically significant difference on time on task between the MCCM-integrated Group 4 and the SAFI-non-integrated Group 1, $t = 2.36, p = .02$, $\eta^2 = .06$, observed power = .64. There was also a statistically significant difference on time on task between the MCCM-integrated Group 4 and the SAFI-integrated Group 2, $t = 2.43, p = .02$, $\eta^2 = .07$, observed power = .67. There was, however, no statistically significant difference on time on task between the MCCM-integrated Group 4 and the MCCM-non-integrated Group 3, $t = 1.18, p = .24$, $\eta^2 = .02$, observed power = .22.
As can be seen in Table 6, the mean differences on mental effort ratings between Group 4 MCCM-integrated and the other three groups were 0.39 points \( (SE = 0.59) \) for Group 1 SAFI-non-integrated, 0.30 points \( (SE = 0.59) \) for Group 2 SAFI-integrated, and -0.31 points \( (SE = 0.58) \). None of these differences were statistically significant. Detailed results can be seen in Table 6.

(4) How do examinees’ attitudes and impressions toward their respective concept map assessment activity differ among the four different treatment groups?

The results from the multiple-choice questions, Questions 1 to 7 and 11 of the Attitudes and Impressions Questionnaire (Appendix 4), have been organized in Table 7 and are listed in the following paragraphs. 41 percent of SAFI-integrated Group 2 felt that the list often obstructed relevant material, thus making the task harder, while only 17 percent of MCCM-integrated, 9 percent of MCCM-non-integrated, and zero percent of SAFI-non-integrated had such perceptions of their respective maps. 39 percent of MCCM-integrated believed that the list obstructed non-relevant information, while 18 percent of SAFI-integrated, 10 percent of SAFI-non-integrated, and 9 percent of MCCM-non-integrated perceived the same about their respective maps. 50 percent of both SAFI-non-integrated and MCCM-non-integrated believed that the list did not obstruct the view of any information on the concept map, but sometimes the distance between the blank and the selection list made it harder for them to make the selection.

As for the preferred location of the list, 65 percent of SAFI-non-integrated, 54 percent of MCCM-non-integrated, 50 percent of SAFI-integrated, and only 30 percent of MCCM-integrated believed that the answers should be on the left or the right side of the
52 percent of MCCM-integrated, 31 percent of both MCCM-non-integrated and SAFI-integrated, and 25 percent of SAFI-non-integrated believed that the list should be within the map exactly on top of a blank.

When looking at how the participants of each group felt about the quantity and quality of alternatives on their respective list, 87 percent of MCCM-integrated thought that the number of potential answers was appropriate, and the task of sifting through them properly challenged their knowledge of the material, whereas only 75 percent of SAFI-non-integrated and 68 percent of both SAFI-integrated and MCCM-non-integrated felt the same about their respective lists. Consequently, only 4 percent of MCCM-integrated believed that the number of potential answers was appropriate, but the task of sifting through unrelated potential answers was time-consuming, distracting, or frustrating, whereas 18 percent of MCCM-non, 15 percent of SAFI-non-integrated, 14 percent of SAFI-integrated felt the same about their respective lists.

A further in-depth look at the participants’ impressions about an appropriate number of alternatives on the concept map selection list, revealed that a great majority of the MCCM groups (82 percent of MCCM-non-integrated and 70 percent of MCCM-integrated), who received lists of 4 alternatives, believed 4 or less to be ideal. However, the majority of the SAFI groups who received lists of 13 alternatives (65 percent of SAFI-non-integrated and 68 percent of SAFI-integrated) believed that the ideal number of alternatives on a selection list is ‘5 to 10’.

The majority of each group also believed that the design was satisfactorily user-friendly: 80 percent of SAFI-non-integrated, 73 percent of SAFI-integrated, 64 percent of MCCM-non-integrated, and 61 percent of MCCM-integrated. When probed further,
however, 20 percent of SAFI-non-integrated reported that the primary cause for their mental effort or strain was the poor assessment design more so than their lack of knowledge of the content matter, versus 9 percent of SAFI-integrated, 4 percent of MCCM-integrated, and zero percent of MCCM-non-integrated.

When asked what slowed them down the most, 27 percent of SAFI-integrated, 15 percent of SAFI-non-integrated, 13 percent of MCCM-non-integrated, and only 4 percent of MCCM-integrated reported that it was the processing of irrelevant potential answers on the selection list.

Sixty-four percent of MCCM-non-integrated, 55 percent of SAFI-integrated, 44 percent of MCCM-integrated, and 35 percent of SAFI-non-integrated responded that they would recommend large-scale use of the concept mapping assessment in which they participated.

Qualitative Results

When analyzing the participant responses collected from the three open-ended questions, Questions 8, 9, and 10 in the Attitudes and Impressions Questionnaire, the data were sorted into categories, then tallied. “Emergent categories” (Creswell, 2007, p. 152) were used, rather than pre-existing codes, since the categories emerged from comparing participants’ comments and grouping them together based on similarities among them. As can be seen in Table 8, the first open-ended question (Question 8) had a total of 9 emergent categories, the second open-ended question (Question 9) had a total of 6 emergent categories, and the third open-ended question (Question 10) had a total of 7 emergent categories. These emergent categories and their respective group percentages
are all listed in Table 8. The next three subsections of this chapter list the more significant findings of each of the three open-ended questions.

**Question 8: In your own words, what did you like most about this type of assessment?**

Thirty percent of SAFI-non-integrated and 27 percent of SAFI-integrated stated that they liked that there was a list to choose from. Some of these participants referred to the selection list as a nice way to prompt their knowledge or to give a hint, which would not be available for recall questions or fill-in-the-blank maps. 18 percent of SAFI-integrated, 20 percent of SAFI-non-integrated, and 23 percent of the MCCM-non-integrated liked the easy/user-friendliness of their respective map design. However, a substantially higher 43 percent of the MCCM-integrated group liked the easy/user-friendliness of their respective map design. Additionally, 13 percent of the MCCM-integrated group liked the integrated placement of the selection list and 26 percent commented on liking the multiple-choice format of the selection list. Table 8 contains detailed results of all emergent categories for this open-ended question.

**Question 9: In your own words, what did you like least about this type of assessment?**

Thirty percent of SAFI-non-integrated reported that they disliked the feature of the assessment that required them to go back and forth between the list and the map and 20 percent of SAFI-non-integrated reported that they disliked the length of the list and how repetitive it became. The least favorite part of SAFI-integrated was the fact that the selection list obstructed view of important parts of the map (28 percent) and that the assessment map was too large (23 percent). 23 percent of MCCM-non-integrated
commented that they did not like that the selection lists disappeared, and 26 percent of MCCM-integrated said that they disliked that the assessment was too large due to either too many bubbles or too many words in the map. Table 8 contains detailed results of all emergent categories for this open-ended question.

Question 10: What are your recommendations to help us improve this interactive computer-based assessment?

Twenty-three percent of SAFI-integrated made recommendations to place the selection list to the side of the map (non-integrated) and to provide a smaller map. 14 percent of MCCM-non-integrated stated that the list should be inside the map (integrated), and 13 percent of MCCM-integrated mentioned that the list should be outside the map. 17 percent of MCCM-integrated also recommended the use of smaller maps. Table 8 contains detailed results of all emergent categories for this open-ended question.
CHAPTER 5

DISCUSSION

Following the Summary of the Study, this chapter aims to provide an in-depth interpretation, analysis, and synthesis of the results/findings of the study, as well as the limitations of the study. The Educational Implications section includes some recommendations about current educational assessment policies and practices. The Future Research section focuses on recommending related areas of research that need closer examination. Finally, a Concluding Statement is provided at the end of this chapter.

Summary of the Study

The need to assess schematic knowledge structures (Pellegrino, Chudowsky, & Glasser, editors, 2001) has driven many educational researchers and psychometricians to investigate alternate forms of assessment such as essay writing, open-ended questions, interviews, think-alouds, concept maps, performance assessments, portfolios, and others. Concept maps quickly became a favorite among alternative assessments primarily because of their graphical/schematic layout appeal for organizing knowledge. According to O’Neil & Klein (1997), concept maps are graphical representations of students’ understanding of interrelated concepts. They contain labeled nodes or bubbles in the shape of circles or rectangles that are connected by labeled links/lines or lined labels. The nodes are labeled by concepts and the connecting links/lines are labeled by the relationships among the concepts. The concept map assessment method that possesses much-desired psychometric properties of high validity and reliability (Chang, Sung, and Chen, 2001; Ruiz-Primo, 2004) is the select-and-fill-in (SAFI) concept map. Nonetheless,
this study has examined two potential areas of improvement in the assessment design of 
traditional SAFI concept maps by evaluating the effects of (1) the type of selection list of 
concepts and relations and (2) the spatial placement of the selection list on the 
examinee’s overall mental effort.

The Traditional SAFI Concept Map

and its Traditional List (TL)

The first potential area of improvement of traditional SAFI concept maps 
pertaining to the type of list of concepts and/or relations to be selected-and-filled-in the 
partially filled expert concept map came from research findings in both cognitive science 
and psychometrics. The traditional list (TL) of concepts and/or relations used by 
traditional SAFI concept maps usually contains about 10 to 20 concepts or relations or a 
combination of both concepts and relations. The assessment task of the student/examinee 
is to select a concept or relation from the list and place/write/type/click it into the 
appropriate blank node or blank link within the concept map; naturally, this task is 
replicated with every attempt to fill in a blank node or blank link. This study aimed to 
examine whether the traditional SAFI task of repetitively sifting through this traditional 
list (TL) does, in fact, impose unnecessary mental strain on the examinee, namely by 
virtue of the task being excessively time-consuming, frustrating, and even cognitively 
overwhelming.

This potential cognitive strain supposedly imposed by the traditional list (TL) has 
direct implications for working memory (Grimley & Banner, 2008; Pickering & 
Gathercole, 2004; Miyake & Shah, 1999; Baddeley, 1986), which is the main information 
processor or the locus of control for our cognitive processes and is limited in both
capacity and duration when handling information (Baddeley, 1986; Miyake & Shah, 1999). Thus, in an attempt to improve the quality of traditional SAFI concept maps, the objective of the study was to use relevant research findings from cognitive science and psychometrics to create and then test a new type of selection list(s) of concepts or relations, the multiple-choice list (MCL).

The Multiple-Choice Concept Map (MCCM) and its Multiple-Choice List (MCL)

Based on research findings from cognitive science dealing with working memory and human cognitive architecture and its limitations (Miller, 1956; Peterson & Peterson, 1959; Baddeley, 1986; Sweller, van Merrienboer, & Paas, 1998; Kalyuga, Ayres, Chandler, & Sweller, 2003), the number of options (in the form of concepts or relations) contained in a selection list have been reduced by excluding all non-related or construct-irrelevant alternatives, so as to avoid unnecessary cognitive processing. Based on psychometric research findings on the optimum number of alternatives/options for multiple-choice questions (Waller, 1989; Hutchinson, 1997; Abad, Olea, & Ponsoda, 2001), the selection list contained no more than four concepts or relations. These four carefully-chosen alternatives (whether concepts or relations) contained one single correct answer and three incorrect-but-plausible alternatives to the corresponding blank node or blank link. Therefore, each blank node and each blank link had a corresponding carefully-tailored Multiple-Choice List (MCL) of four (or fewer) alternatives. The concept map method that makes use of the multiple-choice list (MCL) is called the Multiple-Choice Concept Map (MCCM) because its design borrows from both traditional concept map assessment methods and traditional multiple-choice item assessment.
Both the traditional SAFI concept map and the MCCM used in this study were computer-based and the user interactivity was identical. Namely, the TL and the MCL were invisible by default and made visible through a single (left) click of the mouse. Upon viewing the TL or the corresponding MCL, the student would make his/her selection by another single (left) click of the mouse that would automatically place his/her selection in the corresponding blank node or blank link.

**Traditional (Non-Integrated) Spatial Placement of the Selection List**

The second potential area of improvement of traditional SAFI concept maps pertaining to the spatial placement of the list of concepts and/or relations came out from research findings in cognitive science (Chandler and Sweller, 1991, 1992; Cerpa, Chandler, & Sweller, 1996). Traditional SAFI concept maps, whether used for instructional purposes or assessment purposes (formative or summative), typically present the selection list of concepts or relations separate from the partially completed expert concept map, usually to the side or on top of it. If the traditional SAFI maps are on paper as opposed to on a computer, these lists are often on separate sheets of paper, and, in the case of poor design, even on the backside of the concept map sheet. The limitations of working memory (Baddeley, 1986; Miyake & Shah, 1999) make it difficult for a student to cognitively process the information on the map and the information on the list simultaneously. Thus, the examinee is forced to expend valuable and limited cognitive resources by intentionally directing his/her attention back and forth between the two separate sources of information (the map and the list) in order to complete the assessment task of selecting items from the list and filling in the existing blank nodes and links in the
map. This study aimed to examine whether the traditional SAFI assessment design of non-integrated selection lists does, in fact, unnecessarily increase the student’s total mental effort and overload his/her working memory or cognitive capacity.

The aforementioned phenomenon of increased cognitive load or mental effort attributed to the need to split one’s attention between two sources of information has been extensively researched in the area of instructional design and termed the “split-attention effect” (Chandler and Sweller, 1991, 1992; Cerpa, Chandler, and Sweller, 1996). Thus, in an attempt to improve the quality of traditional SAFI concept maps, maybe even the MCCM, the objective of the study was to use relevant research findings from instructional design to create and then test a computer-based interactive concept map that used an Integrated Selection List that appeared within the concept map and right over the corresponding blank node or blank link.

Non-Traditional (Integrated) Spatial Placement of the Selection List

Based on research findings from the area of instructional design (Chandler and Sweller, 1991, 1992; Cerpa, Chandler, and Sweller, 1996) and, in attempt to determine whether the detrimental consequence of increased mental effort or cognitive load caused by the split-attention effect makes its appearance in traditional (non-integrated) concept map assessment, both the traditional list (TL = 10 to 20 options/alternatives) and the multiple-choice list (MCL = 4 options/alternatives) have been integrated inside the actual concept map. In other words, both the TL and the MCL were spatially presented within the concept map and right over top of the corresponding blank node or blank link, unlike
the traditional or non-integrated TL and MCL which were presented separate from or adjacent to the concept map.

The user interactivity, however, remained identical to the non-integrated concept maps. Namely, the TL and the MCL were invisible by default and made visible (within the concept map) through a single (left) click of the mouse. Upon viewing the TL or the corresponding MCL (spatially placed right over top of the blank node/link), the student would make his/her selection by another single (left) click of the mouse that would automatically place his/her selection in the corresponding blank node or blank link.

**Purpose and Significance of the Study**

This study attempted to bridge the gap between cognitive psychology and educational measurement, which is a necessary step in the quest to advance educational assessment (Mislevy, 2008; Leighton & Gierl, 2007; Nichols, 1994; Messick, 1989; Snow & Lohman, 1989). More specifically, this research investigated what promises to be an improved concept map assessment method at a time when the National Assessment Governing Board (NAGB) of the U.S. Department of Education is recommending the nationwide use of concept mapping for student assessment purposes. In its report entitled *Science Framework for the 2009 National Assessment of Educational Progress* the board recommends the use of Select-And-Fill-In (SAFI) concept maps. This research investigated two potential areas of improvement of traditional SAFI concept map assessment based on hypothesized weaknesses rooted in research finding from cognitive science and psychometrics. More specifically, cognitive theories from working memory (Baddeley, 1986; Miyake & Shah, 1999; Grimley & Banner, 2008), multimedia learning (Mayer, 2001) and cognitive load (Chandler and Sweller, 1991, 1992; Cerpa et al., 1996)
have been used to identify potential design weaknesses of traditional SAFI concept map assessment and then to guide the design of the improved and recommended MCCM assessment method.

**Research Design and Questions**

Since one of the primary objectives of this study was to examine the effects of (1) the type of the selection list (TL vs. MCL) and (2) the spatial placement of the selection list (integrated vs. non-integrated) on the examinee’s overall mental effort, participants were randomly assigned to four different groups (SAFI-non-integrated using TL; SAFI-integrated using TL; MCCM-non-integrated using MCL; MCCM-integrated using MCL). Using a 2x2 Factorial MANCOVA, with prior knowledge of the content matter as the Confounding Variable, participants were compared on time on task and mental effort. The Attitudes and Impressions Questionnaire provided additional data for further analytical insight into answering the following research questions: (1) What is the effect of the type of list/map used (TL/SAFI and MCL/MCCM) on the length of time to complete the assessment and on the examinees’ exerted mental effort? (2) What is the effect of the spatial placement of the selection list (integrated and non-integrated) on the length of time to complete the assessment and on the examinees’ exerted mental effort? (3) Does Group 4 (MCCM-integrated using MCL inside the map) take less time to complete the task and exert less mental effort during the assessment than the three other groups? (4) How do examinees’ attitudes and impressions toward their respective concept map assessment activity differ among the four different treatment groups?
In-Depth Discussion of the Results/Findings

(1) What is the effect of the type of list/map used (TL/SAFI and MCL/MCCM) on the length of time to complete the assessment and on the examinees’ exerted mental effort?

The 2x2 factorial MANCOVA results in Table 4 show that the type of List/Map had a statistically significant main effect on the overall mental effort/strain or the linear combination of time on task and mental effort rating scores. More specifically, however, the test of between-subjects effects indicated that the type of list/map had a statistically significant effect on the time to complete the assessment, but not on the examinees’ mental effort ratings. In other words, participants who received the MCCM assessments (which used the MCL) completed the assessment task significantly faster but did not rate themselves higher or lower on personal mental effort than the participants who received the traditional SAFI concept map assessments.

The fact that the effect of the type of list on the time on task was significant suggests that the use of the MCL minimizes the presumed weaknesses of traditional SAFI concept maps by reducing unnecessary mental strain during the assessment. This is crucial, as increased demands on working memory negatively affect test performance (Grimley & Banner, 2008; Pickering & Gathercole, 2004), rendering inaccurate test scores and, in turn, less valid inferences. Minimizing the processing of content-irrelevant information and doing away with the unnecessary assessment cognitive processes (Mislevy et al., 2010; Baker, 2007; Stout, 2002) enables MCCM assessments to make more of the examinee’s limited cognitive capacity readily available for the much needed cognitive processing of content-relevant information.
MCCMs’ efficient solicitation of the examinee’s limited cognitive capacity is important also because it may decrease the chances that the examinee would become frustrated or cognitively overwhelmed due to poor assessment design, as the case may be during traditional SAFI assessments. This is essential because a frustrated and cognitively overwhelmed examinee will not perform to their potential and, thus, the end result will be assessment scores that measure the examinee’s knowledge inaccurately (Grimley & Banner, 2008; Pickering & Gathercole, 2004). The efficient use of the MCL by the MCCM, on the other hand, solicits the examinee’s cognitive processing resources only for content-relevant information and, therefore, it yields scores that measure the examinee’s knowledge more accurately.

Time on task alone, however, is not an adequate measure of extraneous cognitive load. From a logistics perspective, merely reading or processing the total number of alternatives in the assessment also affects the time on task. The MCCMs used MLCs that contained sets of 4 alternatives that were individually tailored to each of the 17 blanks (nodes and links) and no alternative was used twice if it did not relate to the blank in question. Similarly to alternatives of traditional multiple-choice items, MCL’s 4 alternatives were tailored to test the examinee’s knowledge by containing only one correct answer and three plausible-but-clearly-wrong alternatives. The more traditional SAFI maps used one TL that contained 13 alternatives, which remained unchanged regardless of the blank in question. This TL which was presented every time a blank was to be filled in, contained one correct answer and 12 distracters, most of which were not related to the blank in question because they were the correct answer for other blanks in the map. Thus, the total number of the 17 MCL’s alternatives exceeded the TL’s
unchanging 13 alternatives, yet the MCCM groups completed the same task of filling in the 17 blanks faster. This could be attributed to the fact that the TL’s unchanging 13 alternatives were processed multiple times, potentially 17 times. The difference was that the MCL presented only bits of information relevant to the blank in question at any given time, whereas the TL presented the same bits of information most of which were irrelevant to the blank in question every time a blank was to be filled in. Processing only content-relevant alternatives ensures that the mental effort or cognitive load exerted is intrinsic and necessary for completing the cognitive task of filling in the blank.

Inadvertently, processing content-irrelevant alternatives solicits extraneous cognitive load that can and should be avoided through improved assessment design. Time on task, however, though a valuable measure of overall mental effort (Paas et al., 1994; Paas et al., 2003), does not make a distinction between the type of mental effort it detects. Moreover, even though the design of this study controlled for prior knowledge, other factors (e.g., reading speed) may have affected the time on task, in spite of the fact that the participants were randomly assigned to the different treatments. Thus, supporting data from multiple measures and even more research is required to confirm improvement of traditional SAFI assessment design.

The fact that the effect of the type of list on mental effort ratings was not significant may be attributed to inaccurate mental effort rating scores. The subjective one-item 9-point Likert-type mental effort rating scale used in this study (see Figure 13) was highly recommended by Kalyuga, Chandler, and Sweller (2001) for its high correlation to objective measures (.80 to .99) as well as its non-intrusive nature and by Paas and colleagues (2003) who described this instrument as “valid, reliable, and
unintrusive” (p. 66). However, in addition to potential self-reporting inaccuracies associated with subjective rating scales, the wording used with the scale might have been too verbose (see Figure 13) causing some participant confusion and misguided mental effort evaluation. Alternatively, participants may have rated mental effort attributed to the map content which is associated with intrinsic cognitive load, instead rating their mental effort attributed to the map design which is associated with extraneous cognitive load. Anticipating some of these potential limitations of the mental effort scale, this research used the Attitudes and Impressions Questionnaire (Appendix 4) to further examine participant mental effort attributed to concept map assessment design as opposed to mental effort attributed to concept map content.

An alternative explanation for the type of list having no significant effect on the mental effort rating scores is that the mental effort or cognitive load induced by the cognitive processing of the two types of lists does not differ because the sequential processing of each alternative on the list safeguards the examinee from the cognitive overload that can be caused by the simultaneous processing of multiple pieces of information. Early working memory research findings showed that an individual can remember a series of approximately seven random numbers (Miller, 1956) for a mere several seconds (Peterson & Peterson, 1959), unless the numbers are intentionally rehearsed. Processing seven numbers is done sequentially, then maybe grouped in chunks to aid retrieval. Similarly, sequentially processing 13 TL alternatives or 4 MCL alternatives, which are pieces of information that range from simple facts to actual concepts and relations, can take more than several seconds. An examinee with a high level of prior knowledge of the content matter will be able to eliminate, or even avoid
thoroughly processing, some distracters more efficiently than an examinee with less prior knowledge of the content matter. Moreover, sequentially processing both types of alternatives in an attempt to first eliminate the obvious distracters and then compare the seemingly plausible alternatives is indeed a wise test-taking strategy regardless of expertise level. Naturally, an examinee with a high level of prior knowledge of the content matter, usually identifies and eliminates more obvious distracters than a novice, and maybe more quickly too. When comparing the remaining and seemingly plausible alternatives of a TL or MCL on several relevant factors, however, may have the potential to burden working memory and this could be the reason why research on multiple-choice items found 4 alternatives to be the optimal number (Waller, 1989; Hutchinson, 1997; Abad, Olea, & Ponsoda, 2001). Thus, the number of seemingly plausible alternatives that remains after the process of initial distracter elimination may affect the intensity of the mental strain imposed on the learner. But what may seem to be a plausible alternative to a novice may, in fact, be a very obvious content-irrelevant distracter to an expert. The wording in the mental effort rating scale used in this study prompted the participants to rate “ONLY [the] mental effort caused by the MAP DESIGN user-friendliness or flaws” (see Figure 13) often associated with extraneous cognitive load (Chandler and Sweller, 1991, 1992). The inadequacy of time on task as a single measure of extraneous cognitive load and potential deficiencies of the mental effort rating scale described in this subsection, further confirms the need for more research in this area.
(2) What is the effect of the spatial placement of the selection list (integrated and non-integrated) on the length of time to complete the assessment and on the examinees’ exerted mental effort?

The 2x2 factorial MANCOVA results in Table 5 show that the spatial placement of the selection list (integrated and non-integrated) did not have a statistically significant main effect on the linear combination of time on task and mental effort rating scores. The test of between-subjects effects indicated that the spatial placement of the selection list (integrated versus non-integrated) did not have a statistically significant effect on the time to complete the assessment or on the examinees’ exerted mental effort. In other words, participants who received the integrated lists did not complete the assessment task faster and did not rate themselves lower on personal mental effort than the participants who received non-integrated selection list.

Of noteworthy importance, however, is the fact that when participants were further prompted for their attitudes and impressions about the spatial placement of the list and its effect on their performance on/during the assessment task, 50 percent of both the SAFI-non-integrated participants and the MCCM-non-integrated participants believed that even though the list did not obstruct the view of any information on the actual concept map, it was often harder to make the selection because of the distance between the blank node or link and the selection list. Moreover, it is also possible that the potential detrimental impact of the split-attention effect might have been diminished by the proximity of both non-integrated selection lists (TL and MCL), which were presented on the same computer screen and immediately adjacent to the actual concept map (see Figure 5 and 6). According to Mayer’s spatial contiguity principle (2001) described in
Chapter 2, visual material is best held and processed in working memory when it is presented in close proximity on the same page or screen. Thus, Mayer’s research findings (2001) and the results of this study suggest that the detrimental effect of Chandler and Sweller’s split-attention effect (1991, 1992) makes its appearance only when the physical distance between the two sources of information is large enough.

This unexpected finding that the spatial placement of the selection list might be trivial is further discussed in the section on Research Question (4), which covers the participant attitudes and impressions about the different assessment methods, and in the section on Future Research.

(3) Does Group 4 (MCCM-integrated using MCL inside the map) take less time to complete the task and exert less mental effort while completing it?

Different from an interaction effect about which there was no hypothesis, this research question aimed to evaluate mean differences between the primary group of interest (MCCM-integrated Group 4) and each of the other three groups on both dependent variables (time on task and mental effort) in an attempt to determine whether or not the MCCM integrated Group 4 spent less time on task and exerted less mental effort than the other three groups. The Simple Planned Comparison results (see Table 6) indicate that there was a significant difference on time on task between the MCCM-integrated Group 4 and the SAFI-non-integrated Group 1 as well as the SAFI-integrated Group 2, but not between the MCCM-integrated Group 4 and the MCCM-non-integrated Group 3. In other words, the estimated mean time on task of the MCCM-integrated Group 4 was significantly shorter than that of both SAFI groups (integrated and non-
integrated), but was not significantly shorter when compared to the time on task of the MCCM-non-integrated Group 3. The finding that the MCCM-integrated Group 4, which used the MCL, spent less time on task than both SAFI groups, which used the TL, is in agreement with the results from first research question that the type of list used affects time on task. However, since time on task alone is not an adequate measure of mental effort, data from multiple supporting measures and additional research is necessary to prove that the MCL decreases mental effort. The finding that the MCCM-integrated Group 4, which used the integrated MCL, did not spend less time on task than the MCCM-non-integrated Group 3, which used the non-integrated MCL, further confirms that the placement of the selection list does not affect time on task. In other words, even though time on task alone is not an adequate measure of mental effort, seeing that placing the MCL inside the map (Group 4 treatment) versus leaving it outside of map (Group 3 treatment) has no effect on time on task, further confirms results from the second research question, namely that the spatial placement of the list did not have a significant effect on time on task. Subsequent sections of this chapter will describe a potential reason for why the spatial placement of the selection list had no effect on time on task or on mental effort rating.

The hypothesis of this research question (3) was that MCCM-integrated Group 4 would not only take less time on task than all other three groups (not just Group 1 and Group 2) but also rate lower on mental effort than all other three groups. Similarly and in agreement with this study’s findings from research questions 1 and 2, no statistically significant differences were apparent between MCCM-integrated Group 4 and any of the other three groups on rated mental effort scores (see Table 6). Thus, the alternative
explanations provided in the previous two sections (research questions 1 and 2) serve to reveal potential reasons for this result as well.

(4) How do examinees’ attitudes and impressions toward their respective concept map assessment activity differ among the four different treatment groups?

The participants’ attitudes and impressions, collected using the Attitudes and Impressions Questionnaire (Appendix 4), were meant to provide more insight into examinees’ perceptions (likes and dislikes) about their respective assessment design, as well as some direction into future research. Comparative percentages outlined in Chapter 4 and participant answers to the open-ended questions do, in part, provide supportive feedback to the results of the 2x2 Factorial MANCOVA, but they also reveal some ambiguous findings that raise some questions worth addressing. Thus, this subsection of the chapter will aim to categorize and discuss these results accordingly.

As the results from the 2x2 Factorial MANCOVA indicated about research question 1, participants spent less time processing the multiple-choice lists whether the lists were integrated (inside map) or non-integrated (next to map) and this hypothesized finding is reassuring as it is encouraging to know that the participants’ mental effort was decreased as a result of the use of the MCL (shortened but tailored list) versus the TL (more traditional long list). However, the fact that participants did not rate themselves to have exerted less mental effort as a result of the use of the improved multiple-choice list (MCL) raises some concerns as participant responses to the Attitudes and Impressions Questionnaire may indicate otherwise.
For example, when participants were further prompted about what they thought
the ideal number of alternatives on the selection list should be, very high percentages of
the MCCM groups indicated that their current list contained the ideal number of
alternatives, whereas similarly high percentages of the SAFI groups wanted a shorter
selection list. 20 percent of the SAFI-non-integrated Group 1 (the most traditional SAFI
group) chose to write that they disliked the length of the list even in the open-ended
question about what they like least about their respective assessment. This negative
impression about the lack in quality of the TL and, by extension, the assessment design of
traditional SAFI concept maps, however, does not seem to be in agreement with
participant responses on the mental effort rating scale. These data suggest that the
potential shortcomings of the mental effort scale used in this study might, in fact, be true.
In other words, if a majority of participants who completed the Attitudes and Impressions
Questionnaire have concluded that the assessment design leaves something to be desired,
then one would expect that a similarly large majority of participants who completed the
mental effort rating scale, would rate their exerted mental effort attributed to the
assessment design higher than the participants who were very pleased with the
assessment design they received.

A similarly conflicting finding reveals itself when the participants were asked
about their attitudes and impressions about the quantity and quality of alternatives on
their respective list. The MCCM-integrated Group 4 had the highest percentage of
participants (87 percent) who believed that the number of potential answers was
appropriate and the task of sifting through them properly challenged their knowledge of
the material; it also had the lowest percentage (4 percent = 1 person) of participants who
believed that the number of potential answers was appropriate, but the task of sifting through unrelated potential answers was time-consuming, distracting, or frustrating. Such a result might indicate that the integrated MCL used by the MCCM-integrated Group 4 would impose least overall mental effort or strain on the examinee. The opposite, however, might be true for the more traditional SAFI assessment as suggested by the following quote from a participant in Group 1: “I did not like the number of choices that were on the list. At times it seems like it took me longer to go through the list and find the answer. I would then have to go through and re-read the question. I feel that if the word selection was in the blanks it would be easier.” While one of the dependent variables (time on task) confirms the hypothesis that the MCCM-integrated Group 4 would experience least overall mental effort or strain, the results from research question 3 indicate that the self-rated mental effort scores do not depict that.

As the results from the 2x2 Factorial MANCOVA indicated about research question 2, integrating the selection list (TL and MCL) into the concept map showed no decrease in overall mental effort/strain as a result of the hypothesized reduction or elimination of the detrimental split-attention effect. Contrary to hypothesized results, these findings suggest that the spatial placement of the selection list has no effect on the examinee’s mental effort. In other words, assuming this were true, one would conclude that both the integrated and non-integrated selection lists would solicit equal or similar mental effort attributed to the assessment design. This finding can be simply justified or explained by the fact that the actual spatial or physical distance between the blank nodes/links and the non-integrated selection lists was too small to induce a split-attention effect. As mentioned in the subsection on research question 2, such explanation is based
on Mayer’s spatial contiguity principle (2001) which states that visual material is best held and processed in working memory when it is presented in close proximity on the same page or screen. The non-integrated concept maps (SAFI and MCCM) used in this study were only quasi-traditional with respect to the spatial placement of the selection list. For example, traditional paper-and-pencil SAFI concept maps often provide the selection list on a separate sheet of paper, whereas this study provided the selection list immediately adjacent to the actual concept map.

In addition, participants expressed not only a preference for the MCCM over the more traditional SAFI concept map, but also comparative impressions between the two types of MCCM. In some instances, the examinees seemed to prefer the non-integrated MCCM over the integrated MCCM. As expected though, both the non-integrated MCCM and the non-integrated SAFI concept maps fared well when it came to the selection list potentially obstructing relevant or irrelevant information within the concept maps.

Further inquiry into the attitudes and impressions of the participants, however, generated some results that suggest that the examinees preferred the integrated selection list over the non-integrated selection list and participant comments to open-ended questions even indicate that some were even bothered by the non-integrated assessment design. For example, 50 percent of the participants in the SAFI-non-integrated and the MCCM-non-integrated groups believed that, even though the selection list did not obstruct their view of any information on the concept map, sometimes the distance between the blank node or link and the selection list made it harder for them to make their selection. Thus, half of the participants who received the non-integrated assessments (SAFI and MCCM) stated during the Attitudes and Impressions Questionnaire that the
distance between the selection list and the blank node/link impeded their performance, yet participant personal ratings of exerted mental effort does not confirm the same. In other words, a significant portion of participants indicated that the spatial proximity of Mayer’s spatial contiguity principle (2001) was violated; thus, causing Chandler and Sweller’s split-attention effect (1991, 1992) to unnecessarily burden participant working memory and compromise performance.

Furthermore, participant responses to the open-ended question 9 on what they liked least about their respective assessment indicated that 30 percent of the SAFI-non-integrated group disliked the feature of the assessment that required them to go back and forth between the list and the concept map and 20 percent of SAFI-non-integrated group reported that they disliked the length of the list. To have 30 and 20 percent of participants from the most traditional assessment treatment (SAFI-non-integrated Group 1) identify in an open ended format the two dependent factors of the research design as the very things they liked least about the assessment is very reassuring that this research was well worth conducting and worthy of future investigation. These modest findings confirm the documented need to bridge the gap between educational assessment and cognitive science (Mislevy, 2008; Leighton & Gierl, 2007; Nichols, 1994; Messick, 1989; Snow & Lohman, 1989). More specifically, cognitive theories from working memory (Baddeley, 1986; Miyake & Shah, 1999; Grimley & Banner, 2008), multimedia learning (Mayer, 2001) and cognitive load (Chandler and Sweller, 1991, 1992; Cerpa et al., 1996) do, in fact, reveal these potential concerns with the length of the concept map selection list and its placement relative to the map. The reality of the existence of these two weaknesses of traditional SAFI concept maps, is explicitly summarized by a participant from the SAFI-
non-integrated Group 1 who said that “[t]he amounts of the repeating answer for the blank was very distracting. When I had to look back and forth between the list and the node I was on, I sometimes lost my train of thought since I had to pass over the other bubbles.” This most traditional type of concept map assessment, the non-integrated SAFI concept map, which seemed to be the least favorite in this study, imposes cognitive burden on the examinee, because the aforementioned existing research findings from cognitive science and psychometrics were not used to inform its design.

The SAFI-integrated group where the long TL appeared within the map was the only group who did not like the placement of their respective list; this, of course was to be expected given that the long TL obstructed much of the map whenever the list was visible (see Figure 8). It could also be attributed to the fact that neither integrated group had exposure to the non-integrated group and vice versa. Another speculation, based on numerous participant comments of admiration of and best wishes to the researcher, is that many participants might have chosen to not criticize the assessment so as to not hinder the success of the research/researcher.

Further probing into more general participant attitudes and impressions about the user-friendliness of the different assessment methods revealed that similarly high percentages (ranging from 60 to 80) of participants found the design to be “satisfactorily user-friendly”. When looking at answers to the open-ended question on the least favorite thing about the different assessments, 23 percent of the MCCM-non-integrated participants chose to comment on the user-friendliness as the thing they liked most about the assessment. This percentage was even higher (35 percent) for the MCCM-integrated group who chose to speak about how easy it was to understand and use the map. 26
percent of the MCCM-integrated participants also specified that they liked the non-
conventional multiple-choice format of the concept map selection list. This finding that
the intended user of the MCCM actually liked the format of the assessment is to be
expected in light of the fact that the MCCM really does resemble something familiar to
examinee – the multiple-choice format. The SAFI-non-integrated group, on the other
hand, had a disproportionately higher percentage of participants who reported that the
primary cause for their mental effort or strain was the poor assessment design more so
than their lack of knowledge of the content matter. Additionally and in agreement with
earlier findings of this research, the two SAFI groups had higher percentages of
participants who felt that it was the processing of irrelevant potential answers on the
selection list that slowed them down the most. It is then without surprise that only 35
percent of SAFI-non-integrated participants recommended their assessment method be
used for large-scale assessment as opposed to 63.6 percent of the MCCM-non-integrated
participants.

Limitations of the Study

The 87 participants were students (60 females and 26 males; 1 did not report
gender) from two mid-western colleges who were an average of 22.62 years old and 84
percent Caucasian. Therefore, the findings of this study can be generalized only to a
limited portion of the larger population. Nonetheless, the participant base used for this
study is relatively common for much research conducted on students attending post-
secondary institutions and the findings well-worth noting.
The study was conducted in computer labs under direct supervision of the research coordinator, which made for a positive controlled environment and a smooth process of data collection. However, since all concept maps were computer-based and interactive, all participants had to have at least a minimal comfort level with computers (particularly, browsing the Internet) and concept maps. To address this limitation, all participants had to answer a multiple-choice question regarding their comfort level with browsing the Internet and they all replied that they are at least “somewhat comfortable” or better. In addition, all participants had two minutes to play with or complete a simple practice interactive concept map identical in interactivity with their corresponding Concept Map treatment that followed.

The concept map on the topic of Correlation used in this study (see Figure 1) may have been a little large or busy and, therefore, maybe intimidating to those who had very little or no knowledge about the content matter. Once past the first glance impressions, the participants were able to navigate through the map and fill in the blank nodes and links relatively fast, as the average time on task was 4.53 minutes (see Table 1).

As described in the in-depth discussion on the results of research question 1, the one-item 9-point Likert-type mental effort rating scale used in this study may have posed a limitation to the study due to the wording used in the instructions on how to use it (see Figure 13). This, in turn, may have generated inaccurate scores on the mental effort ratings. In addition and as mentioned earlier, the non-integrated concept maps (SAFI and MCCM) used in this study were only quasi-traditional because the non-integrated selection lists were presented in a much closer proximity to the actual map when compared to traditional SAFI maps. This, as explained by the spatial contiguity principle,
may have been the primary cause for the lack of difference in mental effort between the integrated and non-integrated groups.

Educational Implications

Properly designed concept map assessments, such as the MCCM, can be external knowledge representations with much-desired characteristics of being valid, efficient, and explicit (Mislevy, Behrens, Bennett, DeMark, Frezzo, Levy, Robinson, Rutstein, Stanley, Winters, & Shute, 2007) when assessing students’ knowledge structures or schema-based knowledge (Pellegrino et al., 2001).

The first of three above-mentioned characteristics, namely the psychometric property of high validity (and reliability) of SAFI concept maps, has been confirmed through ample research (Rice et al., 1998; Schau et al., 2001; Klein et al., 2002; Yin & Shavelson, 2005, 2008). The MCCM investigated in this study sought to eliminate potential threats to the validity of traditional SAFI concept maps. Even though the current research does not provide conclusive evidence about which of the two MCCM assessments (integrated or non-integrated) is better, it certainly does, in part, prove that the MCCM is superior to the traditional SAFI concept map because its assessment design does not impose unnecessary mental effort on the examinee that normally threatens the assessment’s validity. An in-depth look at the attitudes and impressions of the participants in this study further confirms these hypothesized results. Moreover, the much-desired psychometric properties of high validity and reliability (Ruiz-Primo & Shavelson, 1996) of conventional multiple-choice tests are equally shared by the MCCM which came to be conceived partly from research findings from the area of multiple-
choice testing (Waller, 1989; Hutchinson, 1997; Abad, Olea, & Ponsoda, 2001). The
added benefit of the MCCM as compared to conventional multiple-choice assessment is
that it does a better job at detecting the knowledge organization of the examinee and
his/her ability to make connections between concepts. Thus, using MCCM assessments
will enable the examiners and administrators to make more accurate inferences about the
student’s ability to make connections and understand relationships among concepts, a
mainstream skill among expert performers (Pellegrino et al., 2001; Kalyuga et al., 1998).

The second of three previously-mentioned desired characteristics, which has
direct implication for using MCCMs in the Educational System either on a small or large
case, is the efficiency of the MCCM (Mislevy et al., 2007). Fortunately, computer-based
MCCM assessments can be administered and scored quickly, easily and inexpensively.
Similarly to conventional multiple-choice tests, item analysis can be performed on
MCCM assessments so as to ascertain the production of accurate, unbiased test scores
and continued quality improvement of the assessment. In addition, visual corrective
feedback in the form of a fully completed expert map (efficiently distributed as a single
sheet of paper or an email attachment) can be an invaluable formative or summative
corrective resource to the students, which they can use for later reference. In addition,
eliminating the need to process irrelevant material during the assessment can shorten
overall administration time of the assessment. Thus, this time on task decrease confirmed
by the results of this research study also improves MCCM’s efficiency.

The third of three desired characteristics mentioned by Mislevy and colleagues
(2007), which constitutes yet another important educational implication, is the
assessment’s ability to be explicit. The MCCM is explicit not only in the sense that the
examinee can understand the task at hand and be able to navigate through it easily, but also by providing valuable feedback to the examiner about preexisting misconceptions and faulty knowledge structures. An additional advantage that the MCCM has over other types of assessments (including conventional multiple-choice tests) is its low language dependence (Stoddart, 2006). This is significant because high language dependence poses yet another risk on the validity of the assessment (Bailey, 2000; Abedi, Leon, & Mirocha, 2003). The Standards for Educational and Psychological Testing (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 1999) state that assessments that are high on language dependence may introduce construct-irrelevant components into the testing process and, thus, hinder their ability to accurately reflect the competencies which the assessment intended to measure (Boscardin, Jones, Nishimura, Madsen, & Park, 2008). Thus, the ability of the MCCM to be so explicit yet remain the lowest on language dependence (Stoddart, 2006) maximizes its potential to produce accurate, unbiased scores, which is a remarkable trait to possess.

In conclusion, the wide use of the MCCM in the Educational System is encouraged by this research as an alternative or supplement to, not a drastic replacement of, the still valid and reliable conventional multiple-choice assessment. The MCCM’s strong foundation in both cognitive psychology and educational measurement, coupled with its inherent psychometric properties, its high efficiency and low language dependence, make it a key contender for large-scale use in both formative and summative assessments. At a time when such bridges between the two sciences are encouraged (Mislevy, 2008; Leighton & Gierl, 2007; Nichols, 1994; Messick, 1989; Snow &
Lohman, 1989), the MCCM should be well-received not only by researchers and educators, but (according to this research) also by the participating examinees/students who seemed to have a higher preference for them than for the more traditional SAFI concept maps. Students’ familiarity with conventional multiple-choice assessments used for decades with much success and their increased computer literacy can help the task of increasing the usage of the MCCM be quite effortless.

Future Research

As described in earlier sections of this chapter, this researcher speculates that the wording used with the one-item 9-point Likert-type mental effort rating scale might have been verbose enough to cause confusion and even produce scores of intrinsic cognitive load or mental effort attributed to content instead of scores of extraneous cognitive load or mental effort attributed to faulty map design (see Figure 13). Thus, duplication of similar research that improves on the specified potential shortfall would render more accurate mental effort rated scores that might better harmonize with participant attitudes and impressions and, hopefully, other more objective mental effort measures as well. Further examination as to the type of mental effort that is affected by assessment design would be valuable insight into the development of concept map assessment.

Another potential limitation described earlier was the fact that the non-integrated concept maps (SAFI and MCCM) used in this study were only quasi-traditional because the non-integrated selection lists were presented in a much closer proximity to the actual map than the non-integrated selection list of a more traditional SAFI map (e.g., paper-and-pencil concept map where the list of concepts/relations is often on a separate sheet of
paper). This closer proximity, as explained by the spatial contiguity principle, may have been the primary cause for the lack of difference in mental effort between the integrated and non-integrated groups. Since the results from this study about the spatial placement of the selection list are inconclusive (given participant attitudes/impressions and the speculated shortcoming of the mental effort rating scale), it would be wise to further investigate whether or not the detrimental split-attention effect makes its appearance in more traditional SAFI and MCCM assessments where the non-integrated selection list is displayed further away from the map as maybe a more traditional SAFI map would.

An interesting finding of this research and, in this researcher’s opinion, one worth investigating in future research is the fact that similarly high percentages of participants in both integrated groups (23 percent for SAFI-integrated and 26 percent for MCCM-integrated) perceived the concept map to be too big (either too many bubbles or too many words). No such statements were made by either of the non-integrated groups. Maybe improved technologies that can zoom into specific areas of the map so as to not visually overwhelm the examinee would be another wise and potentially intriguing element of future research.

Something that was beyond the scope of this study, but that could be valuable for future research is doing a MCCM item analysis on participant responses for better insight into structural deficiencies of the MCCM. Last but certainly not least, as with the introduction of any new type of assessment method or measure, future research that compares it to other forms of assessment such as conventional multiple-choice tests, essays, etc. is necessary and encouraged so as to establish high levels of validity for the new interactive computer-based MCCM.
Concluding Statement

The Interactive Computer-Based Multiple-Choice Concept Map (MCCM) is a new interactive computer-based assessment method that bridges well-documented research findings from cognitive science and psychometrics and it promises to address previously mentioned concerns of the Committee on the Foundations of Assessment (Pellegrino et al., 2001) and the National Assessment Governing Board of the US Department of Education (NAEP, 2008).
### Table 1

**Demographic Information**

<table>
<thead>
<tr>
<th>Demographic Variable</th>
<th>Number of participants who responded</th>
<th>Percentage of participants in this demographic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender ($n = 86$)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>60</td>
<td>70</td>
</tr>
<tr>
<td>Male</td>
<td>26</td>
<td>30</td>
</tr>
<tr>
<td>Type of College ($n = 86$)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Public</td>
<td>64</td>
<td>74</td>
</tr>
<tr>
<td>Private</td>
<td>22</td>
<td>26</td>
</tr>
<tr>
<td>Undergrad vs. Graduate ($n = 86$)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Undergraduate</td>
<td>83</td>
<td>97</td>
</tr>
<tr>
<td>Graduate</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>Ethnicity ($n = 86$)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Caucasian</td>
<td>73</td>
<td>84</td>
</tr>
<tr>
<td>Ethnic Minority</td>
<td>13</td>
<td>16</td>
</tr>
<tr>
<td>Treatment ($N = 87$)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group 1 (SAFI-non-integrated)</td>
<td>20</td>
<td>23</td>
</tr>
<tr>
<td>Group 2 (SAFI-integrated)</td>
<td>22</td>
<td>25</td>
</tr>
<tr>
<td>Group 3 (MCCM-non-integrated)</td>
<td>22</td>
<td>25</td>
</tr>
<tr>
<td>Group 4 (MCCM-integrated)</td>
<td>23</td>
<td>26</td>
</tr>
</tbody>
</table>
Table 2

*Descriptive Statistics*

<table>
<thead>
<tr>
<th>Condition / Group</th>
<th>n</th>
<th>Mean</th>
<th>Estimated Mean</th>
<th>SE</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Time on Task (DV1)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group 1  SAFI-non-integrated</td>
<td>20</td>
<td>293.05</td>
<td>294.40</td>
<td>17.98</td>
<td>71.59</td>
</tr>
<tr>
<td>Group 2  SAFI-integrated</td>
<td>22</td>
<td>295.09</td>
<td>294.87</td>
<td>17.13</td>
<td>96.53</td>
</tr>
<tr>
<td>Group 3  MCCM-non-integrated</td>
<td>22</td>
<td>267.59</td>
<td>264.97</td>
<td>17.18</td>
<td>95.30</td>
</tr>
<tr>
<td>Group 4  MCCM-integrated</td>
<td>23</td>
<td>235.00</td>
<td>236.55</td>
<td>16.77</td>
<td>57.14</td>
</tr>
<tr>
<td>SAFI (integrated+non-int.)</td>
<td>42</td>
<td>294.12</td>
<td>294.63</td>
<td>12.41</td>
<td>84.55</td>
</tr>
<tr>
<td>MCCM (integrated+non-int.)</td>
<td>45</td>
<td>250.93</td>
<td>250.76</td>
<td>11.98</td>
<td>78.98</td>
</tr>
<tr>
<td>Non-integrated (SAFI+MCCM)</td>
<td>42</td>
<td>279.71</td>
<td>279.69</td>
<td>12.42</td>
<td>84.81</td>
</tr>
<tr>
<td>Integrated (SAFI+MCCM)</td>
<td>45</td>
<td>264.38</td>
<td>265.71</td>
<td>11.98</td>
<td>83.68</td>
</tr>
<tr>
<td>Population (N)</td>
<td>87</td>
<td>271.78</td>
<td></td>
<td></td>
<td>84.09</td>
</tr>
<tr>
<td><strong>Mental Effort Rating (DV2)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group 1  SAFI-non-integrated</td>
<td>20</td>
<td>4.70</td>
<td>4.76</td>
<td>0.44</td>
<td>2.20</td>
</tr>
<tr>
<td>Group 2  SAFI-integrated</td>
<td>22</td>
<td>4.68</td>
<td>4.67</td>
<td>0.42</td>
<td>2.25</td>
</tr>
<tr>
<td>Group 3  MCCM-non-integrated</td>
<td>22</td>
<td>4.18</td>
<td>4.06</td>
<td>0.42</td>
<td>1.65</td>
</tr>
<tr>
<td>Group 4  MCCM-integrated</td>
<td>23</td>
<td>4.30</td>
<td>4.37</td>
<td>0.41</td>
<td>2.20</td>
</tr>
<tr>
<td>SAFI (integrated+non-int.)</td>
<td>42</td>
<td>4.69</td>
<td>4.72</td>
<td>0.30</td>
<td>2.20</td>
</tr>
<tr>
<td>MCCM (integrated+non-int.)</td>
<td>45</td>
<td>4.42</td>
<td>4.22</td>
<td>0.29</td>
<td>1.93</td>
</tr>
<tr>
<td>Non-integrated (SAFI+MCCM)</td>
<td>42</td>
<td>4.43</td>
<td>4.41</td>
<td>0.30</td>
<td>1.92</td>
</tr>
<tr>
<td>Integrated (SAFI+MCCM)</td>
<td>45</td>
<td>4.49</td>
<td>4.52</td>
<td>0.29</td>
<td>2.21</td>
</tr>
<tr>
<td>Population</td>
<td>87</td>
<td>4.46</td>
<td></td>
<td></td>
<td>2.06</td>
</tr>
</tbody>
</table>

*Note.* Estimated Means adjusted for Prior Knowledge Covariate evaluated at the following values: Pretest Score = 4.62 ($SD = 2.33$).
Table 3

Pearson Correlations

<table>
<thead>
<tr>
<th>Variable</th>
<th>Prior Knowledge (Covariate)</th>
<th>Time on Task</th>
<th>Mental Effort Rating</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prior Knowledge (Covariate)</td>
<td>1</td>
<td>-0.21*</td>
<td>-0.37**</td>
</tr>
<tr>
<td>Time on Task</td>
<td></td>
<td>1</td>
<td>0.53**</td>
</tr>
<tr>
<td>Mental Effort Rating</td>
<td></td>
<td></td>
<td>1</td>
</tr>
</tbody>
</table>

*p < .05, two-tailed. **p < .01, two-tailed.
Table 4

2x2 Factorial MANCOVA for the Composite DV of Time on Task and Mental Effort

<table>
<thead>
<tr>
<th>Source / Effect</th>
<th>df</th>
<th>hypothesis, error</th>
<th>F</th>
<th>Partial $\eta^2$</th>
<th>Observed power</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prior Knowledge (Covariate)</td>
<td>2, 81</td>
<td>6.95</td>
<td>.15</td>
<td>.92</td>
<td>.00**</td>
<td></td>
</tr>
<tr>
<td>Type of List (Factor A)</td>
<td>2, 81</td>
<td>3.20</td>
<td>.07</td>
<td>.60</td>
<td>.05*</td>
<td></td>
</tr>
<tr>
<td>List Placement (Factor B)</td>
<td>2, 81</td>
<td>0.61</td>
<td>.01</td>
<td>.15</td>
<td>.54</td>
<td></td>
</tr>
<tr>
<td>Interaction (AxB)</td>
<td>2, 81</td>
<td>0.86</td>
<td>.02</td>
<td>.19</td>
<td>.43</td>
<td></td>
</tr>
</tbody>
</table>

*p < .05. **p < .01.
Table 5

*Between-Subjects Effects on the 2 Dependent Variables (DV*s)*

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>Partial $\eta^2$</th>
<th>Observed power</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Time on Task (DV1)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pre-test (Covariate)</td>
<td>26398.23</td>
<td>1</td>
<td>26398.23</td>
<td>4.09</td>
<td>.05</td>
<td>.52</td>
<td>.05*</td>
</tr>
<tr>
<td>Type of List (A)</td>
<td>41721.28</td>
<td>1</td>
<td>41721.28</td>
<td>6.46</td>
<td>.07</td>
<td>.71</td>
<td>.01**</td>
</tr>
<tr>
<td>List Placement (B)</td>
<td>4232.04</td>
<td>1</td>
<td>4232.04</td>
<td>0.66</td>
<td>.01</td>
<td>.13</td>
<td>.42</td>
</tr>
<tr>
<td>Interaction (AxB)</td>
<td>4493.49</td>
<td>1</td>
<td>4493.49</td>
<td>0.70</td>
<td>.01</td>
<td>.13</td>
<td>.41</td>
</tr>
<tr>
<td>Error</td>
<td>529261.86</td>
<td>82</td>
<td>6454.41</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Mental Effort (DV2)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pre-test (Covariate)</td>
<td>53.06</td>
<td>1</td>
<td>53.06</td>
<td>14.03</td>
<td>.15</td>
<td>.96</td>
<td>.00**</td>
</tr>
<tr>
<td>Type of List (A)</td>
<td>5.36</td>
<td>1</td>
<td>5.36</td>
<td>1.42</td>
<td>.02</td>
<td>.22</td>
<td>.24</td>
</tr>
<tr>
<td>List Placement (B)</td>
<td>0.26</td>
<td>1</td>
<td>0.26</td>
<td>0.07</td>
<td>.00</td>
<td>.06</td>
<td>.79</td>
</tr>
<tr>
<td>Interaction (AxB)</td>
<td>0.86</td>
<td>1</td>
<td>0.86</td>
<td>0.23</td>
<td>.00</td>
<td>.08</td>
<td>.64</td>
</tr>
<tr>
<td>Error</td>
<td>310.06</td>
<td>82</td>
<td>3.78</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*p ≤ .05. **p ≤ .01.
**Table 6**

*Simple Planned Comparison*

<table>
<thead>
<tr>
<th>Group / Parameter</th>
<th>Mean Difference</th>
<th>SE</th>
<th>t</th>
<th>Partial $\eta^2$</th>
<th>Observed power</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Time on Task (DV1)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group 1 SAFI-non-integrated</td>
<td>57.85</td>
<td>24.56</td>
<td>2.36</td>
<td>.06</td>
<td>.64</td>
<td>0.02*</td>
</tr>
<tr>
<td>Group 2 SAFI-integrated</td>
<td>58.32</td>
<td>23.97</td>
<td>2.43</td>
<td>.07</td>
<td>.67</td>
<td>0.02*</td>
</tr>
<tr>
<td>Group 3 MCCM-non-integrated</td>
<td>28.42</td>
<td>24.05</td>
<td>1.18</td>
<td>.02</td>
<td>.22</td>
<td>0.24</td>
</tr>
<tr>
<td>Group 4 MCCM-integrated</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Mental Effort (DV2)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group 1 SAFI-non-integrated</td>
<td>0.39</td>
<td>0.59</td>
<td>0.65</td>
<td>.01</td>
<td>.10</td>
<td>0.52</td>
</tr>
<tr>
<td>Group 2 SAFI-integrated</td>
<td>0.30</td>
<td>0.58</td>
<td>0.51</td>
<td>.00</td>
<td>.08</td>
<td>0.61</td>
</tr>
<tr>
<td>Group 3 MCCM-non-integrated</td>
<td>-0.31</td>
<td>0.58</td>
<td>-0.53</td>
<td>.00</td>
<td>.08</td>
<td>0.60</td>
</tr>
<tr>
<td>Group 4 MCCM-integrated</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Note.* Mean Differences were obtained by subtracting Group 4 Estimated Mean from the Estimated Mean of each of the other 3 groups. Estimated Means adjusted for Prior Knowledge Covariate evaluated at the following values: Pretest Score = 4.62 ($SD=2.33$). *p $\leq$ .05.
Table 7

*Quantitative Data – Percentages for Questions 1-7, 11 from the Attitudes Questionnaire*

<table>
<thead>
<tr>
<th>Selected Alternative</th>
<th>Group 1 SAFI-Non</th>
<th>Group 2 SAFI-Int</th>
<th>Group 3 MCCM-Non</th>
<th>Group 4 MCCM-Int</th>
</tr>
</thead>
<tbody>
<tr>
<td>List <em>often obstructed</em> view of relevant information, which made it harder to make selection.</td>
<td>0</td>
<td>41</td>
<td>9</td>
<td>17</td>
</tr>
<tr>
<td>List <em>often obstructed</em> view of non-relevant information, but did not interfere with ability to make my selection.</td>
<td>10</td>
<td>18</td>
<td>9</td>
<td>39</td>
</tr>
<tr>
<td>List did not obstruct view of <em>any information</em>, but sometimes the distance between the blank and the selection list made it harder for me to make my selection.</td>
<td>50</td>
<td>4</td>
<td>50</td>
<td>0</td>
</tr>
<tr>
<td>Prefer the list on the left or right side of the concept map</td>
<td>65</td>
<td>50</td>
<td>54</td>
<td>30</td>
</tr>
<tr>
<td>Prefer the list within the map on top of “blank” to be answered</td>
<td>25</td>
<td>32</td>
<td>32</td>
<td>52</td>
</tr>
<tr>
<td>Number of potential answers was <em>appropriate</em>, and the task of sifting through them properly challenged knowledge.</td>
<td>75</td>
<td>68</td>
<td>68</td>
<td>87</td>
</tr>
<tr>
<td>Number of potential answers was <em>appropriate</em>, but the task of sifting through unrelated potential answers was time-consuming, distracting, or frustrating.</td>
<td>15</td>
<td>14</td>
<td>18</td>
<td>4</td>
</tr>
<tr>
<td>4 or fewer alternatives ideal</td>
<td>20</td>
<td>23</td>
<td>82</td>
<td>70</td>
</tr>
<tr>
<td>5-10 alternatives ideal</td>
<td>65</td>
<td>68</td>
<td>9</td>
<td>26</td>
</tr>
<tr>
<td>Satisfactory user-friendly design</td>
<td>80</td>
<td>73</td>
<td>64</td>
<td>61</td>
</tr>
<tr>
<td>Poor assessment design alone was cause of mental effort</td>
<td>20</td>
<td>9</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>Slowed down mostly by processing the irrelevant potential answers on the selection list</td>
<td>15</td>
<td>27</td>
<td>14</td>
<td>4</td>
</tr>
<tr>
<td>Recommend respective assessment for large-scale use in schools</td>
<td>35</td>
<td>54</td>
<td>64</td>
<td>44</td>
</tr>
</tbody>
</table>
Table 8

Qualitative Data – Percentages for the Open-Ended Questions 8, 9, and 10

<table>
<thead>
<tr>
<th>Emergent Category</th>
<th>Group 1 SAFI-Non</th>
<th>Group 2 SAFI-Int</th>
<th>Group 3 MCCM-Non</th>
<th>Group 4 MCCM-Int</th>
</tr>
</thead>
</table>

**Question 8: In your own words, what did you like most about this type of assessment?**

- Challenged my knowledge: 5 5 5
- Did not obstruct map/info: 10
- Easy/user-friendly design: 20 18 23 43
- Existence of selection list: 30 27 14
- Fun: 5 5
- Integrated selection list: 9 13
- Multiple-choice format: 9 26
- Organization of information: 5
- Visual: 5 9 9 4

**Question 9: In your own word, what did you like least about his type of assessment?**

- Back & forth / split-attention: 30 14
- Important map/info obstructed: 28
- Integrated selection list: 13
- Map too big: 10 23 18 26
- Selection list disappeared: 23
- Selection list too long/repetitive: 20 18

**Question 10: What are your recommendations to help us improve this interactive computer-based assessment?**

- Integrated selection list: 14
- Mark/highlight used selections: 10
- Non-integrated selection list: 23
- Selection list to not obstruct map: 13
- Selection list to remain visible: 9
- Shorter selection list: 10 9
- Smaller map: 10 23 9 17
**Figure 1.** Concept map on the topic of correlation.

![Concept map on the topic of correlation](http://research.cmsfam.com/mccm/cm_correlation.jpg)
Figure 2. Practice concept map on the topic of cats and dogs.

http://research.cmsfam.com/mccm/cm_catsdogs.jpg
Figure 3. SAFI concept map used by Schau, Mattern, Zeilik, & Teague (2001).

http://research.cmsfam.com/mccm/cm_Schau2001.jpg
Figure 4. SAFI concept map used by Chang, Sung, & Chen (2001).

Figure 5. The Select-and-Fill-in (SAFI) concept map non-integrated - Group 1.

http://research.cmsfam.com/mccm/safi_non.jpg
Figure 6. The Multiple-Choice Concept Map (MCCM) non-integrated - Group 3.

http://research.cmsfam.com/mccm/mccm_non.jpg
Figure 7. The Multiple-Choice Concept Map (MCCM) integrated - Group 4.

http://research.cmsfam.com/mccm/mccm_int.jpg
Figure 8. The Select-and-Fill-in (SAFI) concept map integrated - Group 2.

http://research.cmsfam.com/mccm/safi_int.jpg
Figure 9. Practice SAFI on cats and dogs – non-integrated – Group 1.

http://research.cmsfam.com/mccm/cd_safi_non.jpg
Figure 10. Practice SAFI on cats and dogs – integrated – Group 2.

http://research.cmsfam.com/mccm/cd_safi_int.jpg
Figure 11. Practice MCCM on cats and dogs – non-integrated – Group 3.

http://research.cmsfam.com/mccm/cd_mccm_non.jpg
Figure 12. Practice MCCM on cats and dogs – integrated – Group 4.

http://research.cmsfam.com/mccm/cd_mccm_int.jpg
Figure 13. The mental effort rating scale.

http://research.cmsfam.com/mccm/mental_effort_scale.gif
APPENDIX 1

IRB APPROVALS

UNLV
UNIVERSITY OF NEVADA LAS VEGAS

Social/Behavioral IRB – Exempt Review
Approved as Exempt

DATE: January 19, 2010

TO: Dr. Kendall Hartley, Curriculum and Instruction

FROM: Office for the Protection of Research Subjects

RE: Notification of IRB Action by Dr. Paul Jones, Chair
Protocol Title: The Multiple-Choice Concept Map (MCCM): An Interactive Computer-Based Assessment Method
OPRS# 0912-3308M

This memorandum is notification that the project referenced above has been reviewed by the UNLV Social/Behavioral Institutional Review Board (IRB) as indicated in Federal regulatory statutes 45CFR46.

PLEASE NOTE: Data collection at other sites is approved contingent on receipt of Facility Authorization from each site (i.e. Kirkwood Community College and Mount Mercy College).

The protocol has been reviewed and deemed exempt from IRB review. It is not in need of further review or approval by the IRB.

Any changes to the exempt protocol may cause this project to require a different level of IRB review. Should any changes need to be made, please submit a Modification Form.

If you have questions or require any assistance, please contact the Office for the Protection of Research Subjects at OPRSHumanSubjects@unlv.edu or call 895-2794.
Letter of Authorization to Conduct Research at Facility

Brenda Durosinni, MPA, CIP, CIM - Director
Office for the Protection of Research Subjects
University of Nevada Las Vegas
4505 Maryland Parkway Box 451047
Las Vegas, NV 89154-1047

Subject: Letter of Authorization to Conduct Research at Mount Mercy College in Cedar Rapids, IA.

Dear Ms. Durosinni:

This letter will serve as authorization for the University of Nevada, Las Vegas (“UNLV”) research team, Dr. Kendall Hartley and Ioan Ciprian Sas, to conduct the research project entitled The Multiple-Choice Concept Map (MCCM): An Interactive Computer-Based Assessment Method at the Mount Mercy College in Cedar Rapids, IA.

The Facility acknowledges that it has reviewed the protocol presented by the researcher, as well as the associated risks to the Facility. The Facility accepts the protocol and the associated risks to the Facility, and authorizes the research project to proceed. The research project may be implemented at the Facility upon approval from the UNLV Institutional Review Board.

If we have any concerns or require additional information, we will contact the researcher and/or the UNLV Office for the Protection of Research Subjects.

Sincerely,

[Signature]
Facility’s Authorized Signatory

[Name]
IRB Chair

Printed Name and Title of Authorized Signatory

[Date] 12/4/09
January 22, 2010

Brenda Durosinmi, MPA, CIP, CIM -Director
Office for the Protection of Research Subjects
University of Nevada Las Vegas
4505 Maryland Parkway Box 451047
Las Vegas, NV 89154-1047

Subject: Letter of Authorization to Conduct Research at Kirkwood Community
College in Cedar Rapids and/or Iowa City, IA.

Dear Ms. Durosinmi:

This letter will serve as authorization for the University of Nevada, Las Vegas
("UNLV") research team, Dr. Kendall Hartley and Ioan Ciprian Sas, to conduct the
research project entitled The Multiple-Choice Concept Map (MCCM): An Interactive
Computer-Based Assessment Method at Kirkwood Community College in Cedar
Rapids and/or Iowa City, Iowa.

The Facility acknowledges that it has reviewed the protocol presented by the
researcher, as well as the associated risks to the Facility. The Facility accepts the
protocol and the associated risks to the Facility, and authorizes the research project to
proceed upon rental agreements for the use of the computer labs at Kirkwood
Community College.

If we have any concerns or require additional information, we will contact the
researcher and/or the UNLV Office for the Protection of Research Subjects.

Sincerely,

Bill Lamb, Ph.D.

Bill Lamb, Ph.D.
Vice President of Instruction
Kirkwood Community College
319-398-5509
APPENDIX 2

DEMOGRAPHICS QUESTIONNAIRE

Due to the fact that this questionnaire was administered online, it looked differently to the participants (as can be seen in the image below).

However, the content included here is complete and identical.
Demographics Questionnaire

1. What is your gender?
   a. Male
   b. Female

2. What is your age? __________

3. What university or college are you enrolled in?
   a. Kirkwood Community College
   b. Mount Mercy College
   c. University of Nevada, Las Vegas (UNLV)

4. What is your major? __________

5. What is your grade level?
   a. Freshman
   b. Sophomore
   c. Junior
   d. Senior
   e. Graduate student

6. What is your ethnic background? __________

7. How comfortable are you with browsing the Internet?
   a. Not at all comfortable
   b. Somewhat comfortable
   c. Satisfactorily comfortable
   d. Very comfortable

8. Which type of tests do you prefer?
   a. Online tests
   b. Pen and paper tests

9. Have you ever received formal instruction (e.g.: as part of a course) on the statistical concept named ‘correlation’?
   a. Yes
   b. No

10. How knowledgeable are you about the statistical concept named ‘correlation’?
    a. Not at all knowledgeable
    b. Somewhat knowledgeable
    c. Satisfactorily knowledgeable
    d. Very knowledgeable
APPENDIX 3

8-QUESTION PRE-TEST ON THE TOPIC OF CORRELATION

Due to the fact that this pre-test was administered online, it looked differently to the participants (as can be seen in the image below). However, the content included here is complete and identical. Since the question stem in Question *4 mistakenly contained an error (".05" should have been "\.5") which rendered that alternative inaccurate, Question *4 was excluded from the analysis.
Pre-Test on Correlation

Please select the best answer for each of the following Multiple-Choice Questions.

1. Correlation may be defined as a statistical procedure that compares
   a. two groups of subjects on one variable
   b. one group of subjects on two variables
   c. one group of subjects on two or more variables
   d. two groups of subjects on two or more variables

2. The Correlation Coefficient (r) can range from
   a. -.01 to +.01
   b. -1.0 to +1.0
   c. 0.0 to 100
   d. -2.0 to + 2.0

3. Correlation numbers between -.5 and -1.0 indicate a
   a. moderate to strong inverse relationship
   b. weak inverse relationship
   c. moderate to strong direct relationship
   d. negatively skewed relationship

*4. Correlation numbers between .05 and 1.0 indicate a
   a. moderate to strong direct relationship
   b. weak pattern of relationship
   c. positively skewed relationship
   d. statistically significant difference

5. Zero or weak correlation means that students with low scores on one variable in general have
   a. low scores on the other variable
   b. high scores on the other variable
   c. high or low scores on the other variable
   d. lower scores than students in another group

6. Zero or weak correlation means that students with high scores on one variable in general have
   a. low scores on the other variable
   b. high scores on the other variable
   c. high or low scores on the other variable
   d. higher scores than students in another group
7. Negative (inverse) correlation may be best represented by the following example
   a. students with low GPAs have many shoes, and those with high GPAs have few shoes
   b. students with low GPAs have few shoes, and those with high GPAs have many shoes
   c. students with low GPAs have few shoes, but so do students with high GPAs
   d. neither students with high GPAs or students with low GPAs have many shoes

8. Moderate to strong negative (inverse) correlation means that subjects who tend to receive low scores on one variable
   a. attain high scores on the other variable
   b. achieve low scores on the other variable
   c. have negative scores on the other variable
   d. have lower scores than subjects in another group

9. Moderate to strong positive (direct) correlation means that subjects who tend to receive high scores on one variable
   a. attain high scores on the other variable also
   b. achieve low scores on the other variable
   c. have positive scores on the other variable
   d. outscore subjects in the comparison group
APPENDIX 4

ATTITUDES AND IMPRESSIONS QUESTIONNAIRE

Due to the fact that this questionnaire was administered online, it looked differently to the participants (as can be seen in the image below). However, the content included here is complete and identical. Each question was served individually and was validated.
Attitudes and Impressions Questionnaire

Your feedback and suggestions about the Interactive Computer-Based Concept Map on Correlation will be of great help to the researchers. Thus, please take the time to read each question thoroughly and then answer truthfully. Thank you!

1. Think back to when you were selecting the best answer to fill in a “blank” on the concept map (CM). Which one of the following 4 statements best describes your impression of the location of the selection list of potential answers?
   a. It often obstructed my view of relevant information on the CM, which made it harder for me to make my selection.
   b. It often obstructed my view of NON-relevant information on the CM, but did NOT interfere with my ability to make my selection.
   c. It did NOT obstruct my view of any information on the CM, but sometimes the distance between the blank and the selection list made it harder for me to make my selection.
   d. It did NOT obstruct my view of any information on the CM, and did NOT interfere with my ability to make my selection.

2. In relation to the concept map, what do you think would be the best location for the selection list of potential answers?
   a. On the left OR right side of the concept map.
   b. On the top OR bottom of the concept map.
   c. Anywhere within the concept map.
   d. Within the concept map, exactly on top of the “blank” to be answered.

3. Think back to the times when you were evaluating the potential answers in a selection list as you were attempting to fill in a “blank” on the concept map. Which one of the following statements best describes the quantity and the quality of potential right and wrong answers on the selection list?
   a. The number of potential answers was appropriate, and the task of sifting through them properly challenged my knowledge of the material.
   b. The number of potential answers was appropriate, but the task of sifting through unrelated potential answers was time-consuming, distracting, or frustrating.
   c. The number of potential answers was overwhelming, but the task of sifting through them properly challenged my knowledge of the material.
   d. The number of potential answers was overwhelming, and the task of sifting through unrelated potential answers was time-consuming, distracting, or frustrating.

4. What do you think would be the ideal number of potential answers on a selection list of this type of concept map assessment?
   a. 4 or fewer
   b. 5 to 10
   c. 10 to 15
d. 15 to 20

5. Overall, I believe the design of this computer-based assessment was
   a. not at all user-friendly.
   b. somewhat user-friendly.
   c. satisfactorily user-friendly.
   d. very user-friendly.

6. Think back to the overall mental effort you exerted or the strain you experienced while completing the interactive computer-based concept map on the topic of correlation. What would you say was the primary cause of your mental effort or strain?
   a. The poor assessment design only.
   b. My lack of knowledge of the content matter only.
   c. The poor assessment design more so than my lack of knowledge of the content matter.
   d. My lack of knowledge of the content matter more so than the poor assessment design.

7. Think back to how long it took you to complete the interactive concept map on correlation. What slowed you down the most?
   a. My lack of knowledge of the content matter.
   b. Overanalyzing seemingly feasible potential answers on the selection list.
   c. Processing the irrelevant potential answers on the selection list.
   d. Nothing slowed me down. I finished quite quickly.

8. In your own words, what did you like most about this type of assessment?

9. In your own words, what did you like least about this type of assessment?

10. What are your recommendations to help us improve this interactive computer-based assessment?

11. Would you recommend/encourage large-scale use of the interactive computer-based concept map assessment you just participated in?
    a. Yes, I would.
    b. No, I would not.
    c. Yes, I would if it underwent slight improvements.
    d. Yes, I would if it underwent major improvements.
143

BIBLIOGRAPHY


CA: University of California, National Center for Research on Evaluation, Standards, and Student Teaching (CRESST).


VITA

Graduate College
University of Nevada, Las Vegas

Ioan Ciprian Sas
aka John C. Sas

Degrees:
Bachelor of Theology, 1998
St. Andrew’s College, Winnipeg, Manitoba, Canada

Master of International Business, 1999
Saint Mary’s University of Minnesota, Winona, MN

Publications:


Dissertation Title: The Multiple-Choice Concept Map (MCCM): An Interactive Computer-Based Assessment Method

Dissertation Examination Committee:
   Chairperson: Dr. Kendall Hartley, Ph. D.
   Committee Member: Dr. E. Michael Nussbaum, Ph. D.
   Committee Member: Dr. Kent J. Crippen, Ph. D.
   Graduate Faculty Representative: Dr. Sterling J. Saddler, Ph. D.