2011

El-Capstone Scale Development: An Emergent Literacy Concept Inventory

Janet L. Capps
EL-CAPSTONE SCALE DEVELOPMENT:
AN EMERGENT LITERACY CONCEPT INVENTORY

By

JANET L. CAPPS

A Dissertation submitted to the
School of Library & Information Studies
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Degree Awarded
Spring Semester, 2011
The members of the committee approve the dissertation of Janet L. Capps defended on March 17, 2011.

_____________________________
Kathleen Burnett
Professor Directing Dissertation

_____________________________
Christopher J. Lonigan
University Representative

_____________________________
Gary Burnett
Committee Member

_____________________________
Eliza Dresang
Committee Member

Approved:

_________________________________________________________
Dr. Corinne Jörgensen, Director, School of Library and Information Studies

_________________________________________________________
Dr. Larry Dennis, Dean, College of Communication and Information

The Graduate School has verified and approved the above-named committee members.
I dedicate this to Madison Analine.
ACKNOWLEDGEMENTS

Dissertations and psychometric activities do not happen in isolation. This project could not have been completed without the assistance of many people. First and foremost, my gratitude goes out to the volunteer participants in this study, for without them, there would be no data. My deepest appreciation is extended to the numerous participant recruitment supporters (e.g., Foundation for Early Learning, Connect the Dots, Wisconsin Department of Public Instruction, Center for Early Literacy Learning, FSULibIT, CCBC-Net, the folks affiliated with ELPLP and the Washington Libraries, ALSC-L, LM-Net, Child-Lit, M. Capps, R. Davis, B. Diamant-Cohen, N. Everhart, C. Fiore, S. Ghoting, A. Goldsmith, M. Gross, J. Hill, C. Hinson, L. Justice, C. Koontz, M. Mardis, L. Mon, E. Randeree, M. Stansbury, the numerous early learning faculty from Maine to Washington, and the other innumerable Facebook connections).

My sincere thanks to R. Branciforte, M. Davis, and the other wonderful support staff members from the School of Library & Information Studies and the Learning Systems Institute for shuffling the appropriate forms around to keep me funded, attending conferences, and moving forward. Speaking of funding, I would like to thank C. Jörgensen, K. DeMeester, and my other funders for providing me with the multiple learning opportunities over the years. Thank you P. Doffek and the Goldstein Library staff for keeping my essential library resources available without the accruement of overdue fines. A deep appreciation goes to the many fellow graduate students for their collegiality and friendship. A special acknowledgement goes to H. Hawthorne for the numerous IRT conversations and to C. Bickford for his consulting.

An enormous thanks goes out to my advisor K. Burnett and committee members G. Burnett and E. Dresang for their support and guidance throughout my numerous years in Tallahassee. A special thank you to C. Lonigan for introducing me to emergent literacy research and for his willingness to share his knowledge and expertise.

I gratefully acknowledge the central role A. Appiah, M. D., and his Capital Eye Consultants staff had in maintaining my vision throughout this project. Thank you. To my family and friends that have continually supported my efforts, I am forever grateful. Words fall short to describe the important role my parents have in my life. Thank you for the never-ending encouragement, support, and patience. Finally, I have dedicated this dissertation to my daughter to simplify the excessive heartfelt acknowledgements she deserves as my life inspiration!
# TABLE OF CONTENTS

LIST OF TABLES ........................................................................................................................ vii

LIST OF FIGURES ..................................................................................................................... viii

ABSTRACT .................................................................................................................................... x

CHAPTER 1 INTRODUCTION ................................................................................................... 1
  1.1 Background ............................................................................................................................ 2
    1.1.1 Evolving EL Research ................................................................................................... 3
    1.1.2 Evolving Libraries: EL Connections .............................................................................. 8
  1.2 Importance of the Study ...................................................................................................... 19
  1.3 Research Goals ................................................................................................................... 21
  1.4 Limitations .......................................................................................................................... 21

CHAPTER 2 LITERATURE REVIEW ...................................................................................... 23
  2.1 Concept Inventories ............................................................................................................ 23
    2.1.1 The Catalyst: Force Concept Inventory (FCI) ............................................................. 26
    2.1.2 Influential Exemplars ................................................................................................... 28
      2.1.2.01 Basic Technology Competency for Educators Inventory (BTCEI) ..................... 28
      2.1.2.02 Chemistry Concept Inventory (CCI) ................................................................. 29
      2.1.2.03 Conceptual Survey of Electricity and Magnetism (CSEM) ............................... 30
      2.1.2.04 Genetics Concept Inventories ............................................................................ 31
      2.1.2.05 Geoscience Concept Inventory (GCI) ............................................................... 34
      2.1.2.06 Signals and Systems Concept Inventory (SSCI) ................................................ 35
      2.1.2.07 Statics Concept Inventory (SCI₁) ...................................................................... 36
      2.1.2.08 Statistics Concept Inventory (SCI₂) ................................................................... 38
      2.1.2.09 Teaching Perspectives Inventory (TPI) ............................................................. 39
      2.1.2.10 Thermal and Transport Science Inventory (TTSI) ............................................. 39
  2.2 Reliability and Validity ....................................................................................................... 40
  2.3 Theoretical Considerations ............................................................................................... 42
    2.3.1 Classical Test Theory .................................................................................................. 43
    2.3.2 Item Response Theory ............................................................................................... 45
    2.3.3 Contrasting Classical Test Theory and Item Response Theory ............................... 48
  2.4 Summary ............................................................................................................................. 49

CHAPTER 3 RESEARCH DESIGN ........................................................................................... 53
  3.1 Stage One: Pre-participant .................................................................................................. 54
    3.1.1 Content Definition ....................................................................................................... 54
    3.1.2 Draft Instrument Items ............................................................................................... 56
  3.2 Stage Two: Participant ........................................................................................................ 56
    3.2.1 Cognitive Interviews ................................................................................................. 57
      3.2.1.1 Practitioner Participants ....................................................................................... 58
      3.2.1.2 Procedure ............................................................................................................. 58


LIST OF TABLES

Table 1. Relationship of Research Stages to Task Activity.......................................................... 53

Table 2. Interview Session Packets for use in Nine Sessions....................................................... 73

Table 3. Breakdown of the 54 item pool into 2 forms.................................................................. 82

Table 4. Mplus discrimination (a) and difficulty (b) parameters from the Form BCmr data along with the number of item responses and the percentage correct. ............................................... 96

Table 5. Mplus discrimination (a) and difficulty (b) parameters from the Form LCmr data along with the number of item responses and the percentage correct. ........................................... 98

Table 6. The six common linking item information used to establish transformation coefficients to equate LCmr into the scale of BCmr. ...................................................................................... 99

Table 7. Shown are the new discrimination (a) and difficulty (b) parameters on the reversed coded data for LCmr with the linking items noted by shaded blank records. .............................. 100
LIST OF FIGURES

Figure 1. A basic schematic of the Force Inventory Concept illustrating the six major categories and the central role of student misconception sub-topics in the response options. ......................... 27

Figure 2. A basic schematic of the parent/sibling/child relationship connections that may be created using The Brain5 software. .............................................................. 55

Figure 3. An example of a potential parent/child relationship using The Brain5 software for mapping reading readiness as defined by the NELP 2008 report. ............................... 56

Figure 4. An example of the mapping process of the Developing Early Literacy, NELP 2008 report. .................................................................................................................. 64

Figure 5. An example of the chapter 2 mapping process of Developing Early Literacy, NELP 2008 report. ........................................................................................................ 64

Figure 6. An example of the Developing Early Literacy, NELP 2008 report's chapter 2: Meta Analyses mapping of the predictive relationship between variables and decoding dependent on the age assessment. .......................................................... 65

Figure 7. Predictive, definition, and activity-based items comprise the indicator variable categories. .............................................................................................................. 67

Figure 8. Interview activity format of the statements and supplemental survey questions. ...... 74

Figure 9. Coding of the supplemental survey questions from the interview activity. .............. 75

Figure 10. An example of a coded item statement from one of the interview activities. .......... 76

Figure 11. User interface for http://www.elcapstone.com running the script to counter-balance the two forms. ........................................................................................................... 79

Figure 12. The basic script that alternates the participants between the two forms. ............... 79

Figure 13. User interface for the first page of the instrument. ................................................ 80

Figure 14. First question is a mandatory response option to obtain consent. ....................... 81

Figure 15. A scaling diagram demonstrating a response-centered approach. ....................... 84

Figure 16. An ICC showing the relationship of an attribute to the likelihood of endorsement. 90

Figure 17. An ICC example of a low difficulty item. .............................................................. 90
Figure 18. An ICC example of an item discriminating well across ability (θ). ......................... 91

Figure 19. BCmr eigenvalue plot. .......................................................................................... 93

Figure 20. LCmr eigenvalue plot............................................................................................ 94
ABSTRACT

The primary objective of this work was the scale development of a research-based instrument (EL-Capstone) that may be used to measure EL core-knowledge levels of adults working in an environment with expectations of supporting literacy development of young children. Earlier concept inventory, emergent literacy, and test theory literature informed the framework for this assessment instrument.

Designed with three distinct stages, pre-participant, participant, and post-participant, the processes were summarized by six general tasks: (1) define the content, (2) develop and select the instrument items, (3) informal informant sessions, (4) interview sessions with librarians and early literacy practitioners, (5) instrument administration, and (6) evaluation and final refinement. The entire consent-form participant activities centered on developing the items for inclusion in the EL-Capstone inventory.

During the first task, the *Developing Early Literacy, National Early Literacy Panel (NELP) 2008 Report* was adopted as "EL-Expert" surrogates to bind the epistemological framework to the NELP authors' theoretical underpinnings of EL concepts. The NELP Report represents an extremely dense synthesis of the EL experimental and quasi-experimental research. By the end of the pre-participant stage concept-mapping activity, ninety-seven potential instrument items had been constructed that fell into three categories, i.e., predictiveness, definitions, and activities. Accompanying the refined sixty-two items presented during the interview sessions (n=10) were 3 ranking questions, i.e., strength of each concept, difficulty level, and perceived importance. The interview results informed instrument item reductions and modifications. Two alternating online versions of EL-Capstone collected data responses for 30 days from anonymous adult volunteers (n=824). An item response theory two-parameter logistic model (2-PL) was applied during the post-participant stage to identify each item's difficulty and discrimination parameters. When fully developed, this research-based scale may be used to measure EL-levels of adults working in an environment with expectations of supporting literacy development of young children. The subsequent collected information may be used to inform early literacy curriculum development and to support program evaluation.
CHAPTER 1

INTRODUCTION

The primary objective of this work is the development of a research-based instrument (EL-Capstone) that may be used in future research investigations designed to measure the emergent literacy, core knowledge levels of adult practitioners working in an environment with expectations of supporting literacy development of young children. Brief background information introduces this work and sets the scene to facilitate understanding of the connections between emergent literacy research and public libraries. Questions interjected throughout are included as support stimulus intended to help convey the importance of this study.

The literature review introduces and defines characteristics of assessment instruments identified by the term concept inventory (CI). Past approaches to CI instrument construction reveal variability in instrument designs and in the measurements chosen to demonstrate validity and reliability. Commonly acknowledged as the catalyst for subsequent CI research, the Force Concept Inventory (FCI) is presented in detail before highlighting some of the additional CI instruments. The FCI is frequently referenced in the concept inventory literature and it serves as a baseline instrument when describing methodological deviations. The additional inventories included in the literature review have unique design characteristics thought to be potentially informative during the EL-Capstone construction or analysis phase.

The scope of this dissertation is limited to the design of a valid and reliable instrument. As such, the theoretical framework of two test theories, i.e., classical test theory (CTT) and item response theory (IRT) are presented in the literature review since the ultimate theory selection is informed by the data. These two theories guide the decisions related to format and application of statistical measures.

The third chapter is dedicated to the research design. The instrument development process uses three phases. These phases, identified as pre-participant, participant, and post-participant, all have emerging components that are informed through the instrument’s progression. The earliest phase of the design includes: (1) the content
definition of the emergent literacy concepts and any sub-topics and (2) the initial instrument design. The participant phase has both informal and formal participants. Subject matter and field domain experts serve in an informal capacity as informants or advisors during the construction of the item pool. Formally through the use of consent forms, two groups of participants are involved in two distinct data collection tasks related to the development and testing of the instrument (i.e., library practitioners and anonymous adult populations). The final phase, post-participant, is dedicated to data analysis. This study’s endpoint is the initial construction of a valid instrument. Use of EL-Capstone as an instrument to collect data related to the emergent literacy core-knowledge levels of any specific population is reserved for future research.

1.1 Background

Early literacy researchers have made major advances during the past two decades in identifying factors that are predictive of literacy success. The terms emergent literacy and early literacy are used interchangeably within this work since the two terms are supported in the literature. Without any loss to the concepts, both of these terms may appropriately be substituted by the use of EL; however, it should be noted that while the term early learning found in the literature includes an early literacy component, it was seen as a broader term outside the scope of this study.

A call for classroom reform, stronger gains in reading performance outcomes, and improved teacher instruction contributed to an increase in funded research aimed at identifying the components predictive of future reading success. Section 1.1.1: Evolving EL Research quickly summarizes some of the recent policy initiatives that fueled research within the area of early literacy. Accompanying the policy and funding initiatives was a shift in reading paradigms. How this shift translates into the current perspective of reading is introduced along with the complications surrounding the use of the term research. The highlights of the 2008 National Early Literacy Panel (NELP) report on early literacy research is included as NELP is used to define the scope of EL-Capstone item construction.

Section 1.1.2: Evolving Libraries: EL Connections, describes the connections between libraries and communities with emergent literacy research. There are indicators
that suggest public libraries and professional library associations have a perceived level of importance of EL connections with libraries. This section provides examples of some of these indicators that document that public libraries are (1) supporting early literacy, (2) creating partnerships with early learning organizations, (3) allocating resources for early literacy facilities, personnel, professional development, and materials, (4) conducting EL outreach and training for schools, families, and child care providers, and (5) receiving grant money to support these initiatives. Research associated with the effects of early literacy connections with libraries is needed. The design of an instrument that will measure the emergent literacy, core knowledge levels of adult practitioners working in an environment with expectations of supporting literacy development of young children supports this research need.

### 1.1.1 Evolving EL Research

Early childhood literacy is an area of inter-disciplinary research with multiple perspectives and methodological approaches. Notwithstanding differences, there is a consensus in the literature that literacy contributes to the quality of an individual's life and benefits society as a whole. Children who develop strong reading skills progressively become better readers and subsequently have an opportunity for greater academic success and more career choices. Likewise, children with poor reading skills may fall prey to the Matthew Effect. Stanovich coined this term to describe a framework that relates reading to cognitive development and to explain “mechanisms [that] operate to create rich-get-richer and poor-get-poorer patterns of reading achievement” (1986, p. 360).

National interests in literacy research were ignited in part by the educational policies from the 1980s. The calls for reform had perhaps the greatest impact on the increased attention to literacy intervention programs. Identifying the need to improve reading performance and teacher instruction is the federal level report, *A Nation at Risk* (National Commission on Excellence in Education, 1983). Two years later, *Becoming a Nation of Readers* was released by the Commission on Reading (1985). This document was instrumental in creating funding opportunities for literacy research that included the development of parent involvement in child literacy programs. The National Academy of Science created a special committee to investigate the effectiveness of interventions
addressing at risk-children for reading problems. This 1998 commission’s report on the prevention of reading difficulties in young children helped to identify influential reading activities that were occurring in the preschool years. Additional research began to help identify the components that were predictive of future reading success.

Other influential policies and funding opportunities fueled interest in investigating the early literacy development factors. In 2002, the No Child Left Behind Act (2002) was signed by President Bush. The First Lady, Laura Bush, rallied funding for improving literacy, declaring “The years from the crib to the classroom represent a period of intense language and cognitive growth. Armed with the right information, we can make sure every child learns to read and reads to learn” (2001, July 26, p. ¶ 2).

Accompanying the policy and funding was a shift in the paradigm defining reading. The traditional perspective perceived reading as a subject taught in elementary school to a child with an already well-developed vocabulary. This traditional view is associated with the term reading readiness. Influenced by a behaviorist paradigm, conventional reading literacy is defined as a process that is taught once a child has developed to the point where the child is ready to learn to read. Morphett and Washburne determined the age of 6.5 years was an appropriate marker for when a child that was not reading should be taken to the stage of early reading (1931).

A transitional phase in the early 1960s research moved the prepare-for-reading-readiness skills to younger children. Teaching reading readiness sub-skills in kindergarten grew in popularity (Bloom, 1964). By the early 1980s, researchers were arguing that a child’s emerging literacy behavior was influenced through the home and early life experiences (Morrow, 1989; Sulzby, 1985).

Reading is now widely accepted as an emerging process. The skills developed prior to formal reading instruction have been shown to be influential in literacy development. These skills fall into a theoretical framework known as emergent literacy. This perspective diverges from the 1960s traditional view of reading as a magical school event for children around the age of five, and embraces decoding as a process development originating in early childhood (Lonigan, Burgess, & Anthony, 2000). As described by Whitehurst and Lonigan, “Emergent literacy involves the skills, knowledge, and attitudes that are developmental precursors to conventional forms of reading and
writing” (1998, p. 848). Justice and Pullen define emergent literacy as “the precursory knowledge about reading and writing that children acquire prior to conventional literacy instruction and that they bring to the task of learning to read” (2003, p. 99). Other definitions include the concept of a child’s increasing awareness of print knowledge through joint book reading and sense of story (Roth & Baden, 2001).

Emergent literacy concepts are identified and defined differently throughout the early literacy literature. However, among researchers prescribing to the theory of emergent literacy, there appears to be an agreement that there are no clear boundaries between pre-readers and readers. Early literacy researchers are working to identify the best combination of factors to predict reading success. In addition to differences in early literacy definitions, there are paradigm differences related to how to design studies that support these EL investigations. This may, in part, be related to differences in the definition of the term research itself.

The term research has evolved into a generic keyword used to describe a wide range of methodologies and ideologies. While the appropriateness and the selection of the research design should be guided by the specific questions under investigation, the variability in the quality of the research methods and limitations (i.e., validity and reliability factors) complicates outcome synthesis. These factors potentially contribute to an oversimplification and/or overgeneralization of emergent literacy concepts found in both peer and non-peer reviewed publications. Application of theory to practice involves filtering the plethora of information surrounding the very complex domain of emergent literacy concepts. This is not an easy task for practitioners since even the experts are sometimes drawn into debates fostered by misinformation and media distortions (Foorman, Fletcher, Francis, & Schatschneider, 2000). Interdisciplinary reading research institutes, e.g., The Florida Center for Reading Research (FCRR), are collaborating to address these complex issues and to promote the best research-based practices related to literacy instruction. These differences complicate the dissemination of research. National efforts to synthesis the research are apparent in reading literacy reports.

At the turn of this century, the National Institute of Child Health and Human Development (NICHHD) released an evidenced-based assessment of the scientific research-based literature related to teaching reading. Part of the charge was to eliminate
misinformation by defining reading instruction and by clarifying how the research translated into best practice (National Reading Panel, 1999). The Report of the National Reading Panel (NRP): *Teaching Children to Read* (National Institute of Child Health and Human Development, 2000) facilitated changes in policy and educational instruction. However, there were still gaps in the synthesis of research related to early literacy, which is defined as literacy development of children from birth through age five. The newly created National Early Literacy Panel (NELP) was charged with conducting empirical research to discover the factors that support early literacy development and the influences of the home and family on this development. NELP (2008b) focused their attention on answering four questions:

1. What are the skills and abilities of young children (age birth through five years or kindergarten) that predict later reading, writing, or spelling outcomes?
2. Which programs, interventions, and other instructional approaches or procedures have contributed to or inhibited gains in children’s skills and abilities that are linked to later outcomes in reading, writing, or spelling?
3. What environments and settings have contributed to or inhibited gains in children’s skills and abilities that are linked to later outcomes in reading, writing, or spelling?
4. What child characteristics have contributed to or inhibited gains in children’s skills and abilities that are linked to later outcomes in reading, writing, or spelling? (p. 2)

The release of the NELP full report confirmed that a plethora of articles on early literacy have been published. Queries during the NELP investigation resulted in more than 8,000 articles being retrieved from only two databases, the PsychINFO and the Education Resources Information Center (ERIC). Subsequently, these results were narrowed to approximately 500 of the most relevant, robust studies (National Early Literacy Panel, 2008b). Although NELP reported a shortage in scientific studies that could demonstrate correlational evidence between early literacy skills and later predictive literacy development, they did identify six variables (2008a):
• alphabet knowledge (AK): knowledge of the names and sounds associated with printed letters;
• phonological awareness (PA): the ability to detect, manipulate, or analyze the auditory aspects of spoken language (including the ability to distinguish or segment words, syllables, or phonemes), independent of meaning;
• rapid automatic naming (RAN) of letters or digits: the ability to rapidly name a sequence of random letters or digits;
• RAN of objects or colors: the ability to rapidly name a sequence of repeating random sets of pictures of objects (e.g., “car,” “tree,” “house,” “man”) or colors;
• writing or writing name: the ability to write letters in isolation on request or to write one’s own name; and
• phonological memory: the ability to remember spoken information for a short period of time. (p.3)

Lonigan and Shanahan, in the Executive Summary, outlined five additional early literacy skills identified by NELP that were “moderately correlated with at least one measure of later literacy achievement but either did not maintain this predictive power when other important contextual variables were accounted for or have not yet been evaluated by researchers in this way” (2008b, p. 3). The Panel released an additional list of the five potentially important variables as (National Early Literacy Panel, 2008b):
• concepts about print: knowledge of print conventions (e.g., left–right, front–back) and concepts (book cover, author, text);
• print knowledge: a combination of elements of AK [Alphabet Knowledge], concepts about print, and early decoding;
• reading readiness: usually a combination of AK, concepts of print, vocabulary, memory and PA [Phonological Awareness];
• oral language: the ability to produce or comprehend spoken language, including vocabulary and grammar; and
• visual processing: the ability to match or discriminate visually presented symbols. (p. 4)
Additional research is still needed to determine the strongest combination of variables that produce the greatest predictive power on long-term literacy outcomes; however, the NELP report will help to unify the language and focus of early literacy research. The NELP report may be perceived as an attempt by the early literacy research community to conduct an inventory of the research. It should be noted that the NELP Report research methodology selection criteria were similar to the rigorous standards set in the earlier NRP Report. The NELP high standards of including only experimental and quasi-experimental designs substantially reduced the number of studies that were included in the analysis.

As with the earlier NRP Report, some of the NELP Report commentaries reflect differences among researchers in the theoretical approaches and the appropriate statistical methods that are acceptable for supporting early literacy evidenced-based effects as reflected in the post-report commentaries and analyses (Lonigan & Shanahan, 2010; Neuman, 2010; Paris & Luo, 2010; Schatschneider & Lonigan, 2010; Schickedanz & McGee, 2010). Early Literacy is still an emerging inter-disciplinary domain that is accompanied by debate (Pressley, Duke, & Boling, 2004; Schuele & Boudreau, 2008). Commentaries on the NELP report reflect an understanding that more research is still needed in the area of early literacy (Dickinson, Golinkoff, & Hirsh-Pasek, 2010; Gutiérrez, Zepeda, & Castro, 2010; Pearson & Hiebert, 2010; Teale, Hoffman, & Paciga, 2010). For example, Dail and Payne "argue that the findings from the National Early Literacy Panel (NELP; 2008) report related to parent involvement and family literacy programs require further clarification" (2010, p. abstract).

1.1.2 Evolving Libraries: EL Connections

This section provides a look at public library connections to early literacy. Within the peer-reviewed literature, there is an absence of experimental or quasi-experimental studies conducted on the impact library connections have on children’s EL development and long-term conventional literacy gains. There is also a void in this literature that quantifiably measures what early literacy knowledge is held by adults working in library environments with expectations of developing EL skills in young children. What does exist, however, are indicators that document public libraries are (1) supporting early
literacy, (2) creating partnerships with early learning organizations, (3) allocating resources for early literacy facilities, personnel, professional development, and materials, (4) conducting EL outreach and training for schools, families, and child care providers, and (5) receiving grant money to support these initiatives. These indicators suggest that public libraries and professional library associations have a perceived level of importance or value placed on EL connections. Deemed to be important by librarians, these EL connections subsequently translate to be important connections that warrant robust research investigations. It should be noted that the examples included in this section are by no means meant to be comprehensive; however, they do set the stage for understanding the need to conduct research in this area. The importance of EL connections also supports the need for an evidence-based inventory tool that may be used in environments with expectations of nurturing early literacy development in young children.

Libraries have historically made vital connections between children, books, and literacy. Strategically placed within communities, libraries have been well suited for the triangulation of child, parent, and community. Storytimes and summer reading programs have deeply rooted connections with library program offerings and are by no means a new library service to communities. However, for approximately a dozen years, changes have been introduced to children services in public libraries; e.g., physical spaces allocated to children departments and enriched EL storytimes. It is apparent that the incorporation of early literacy research into storytimes appears to have obtained a level of importance in public library policy decisions.

One of these observable changes is the introduction of early literacy components into the traditional programming format as evident through local, state, and national level allocation of resources to support emergent literacy training programs for public libraries. *Every Child Ready to Read@your library* (ECRR) is an example of an evidenced-based EL training program used to enrich classic storytimes with EL components.

The early literacy project ECRR, a joint project of the Public Library Association (PLA) and the Association for Library Services to Children (ALSC) in partnership with the National Institute of Child Health and Human Development (NICHD), utilized this well-established library community placement to access parents and caregivers of young
children (Every Child Ready to Read Evaluation Task Force, 2009). This initiative began in 2000 after the release of the National Reading Panel's report, *Teaching Children to Read: An Evidence-Based Assessment of the Scientific Research Literature on Reading and Its Implications for Reading Instruction*, to assist in the dissemination of emergent literacy findings.

In subsequent years, ECRR expanded with the testing of a researched-based parent-caregiver program model developed by emergent literacy experts Whitehurst and Lonigan (Meyers & Henderson, 2004). The 2002-2003 PLA and ALSC’s impact study measured effects of the ECRR model (for details see the PLA/ALSC *Early Literacy Initiative: 2003 Evaluation* report). The fourteen public libraries involved in this study were literally spread coast to coast across America. This study may have contributed to the increased number of libraries initiating research-based emergent literacy programs since members of the ECRR committee who worked with Drs. Whitehurst and Lonigan were invited to stay on as trainers. Traveling across the country, these six trainers worked closely with both statewide programs and individual systems.

ECRR continues to be an early literacy training framework for public libraries. The ECRR training information has not remained static and the 2009 ECRR Evaluation Task Force included an ECRR revision as one of their seven recommendations (Every Child Ready to Read Evaluation Task Force, 2009). A detailed comparison of these ECRR versions was not available at the time of this study; however, ECRR's call for a revision may be seen as continued support for having early literacy connections to libraries. While specific information about these ECRR changes is unknown, it can be assumed that any changes in training may potentially impact emergent literacy core knowledge levels of practitioners and thus add additional credence to the need for EL-Capstone. It should be noted that in addition to ECRR, other competitive programs have been implemented in public libraries; e.g., Diamant-Cohen's Mother Goose on the Loose, or Reach Out and Read. Having the availability of multiple training venues broadens the scope of agencies perceiving the importance of EL library connections. This variability across programs also suggests that there could be additional possibilities for practitioners to adopt different approaches to defining how early literacy is incorporated into library program services.
The trend of emergent literacy, learning environment partnerships are continuing to develop between libraries and healthcare organizations, libraries and schools, libraries and day care centers, and libraries and communities. These cooperative collaborations may help achieve increases in literacy opportunities for young children; however, at this time, a dearth of outcome-based studies conducted to measure the learning outcomes of the children participating in these library outreach programs exists. For an insightful view into the depth of some of the existing partnerships and commitments, see the whitepaper titled *The Early Literacy Landscape for Public Libraries and Their Partners* (Dresang, Burnett, Capps, & Feldman, 2011). Existing partnerships are used to demonstrate a perceived level of importance placed on EL. Variability in these partnerships serves as an additional indicator that variance may also be occurring in the training and early literacy knowledge levels of practitioners. Conducting research to understand if these differences impact the level of early literacy development of children attending library storytimes may be tied to the level of EL knowledge held by the adults leading these storytimes. The development of EL-Capstone supports future research endeavors in these areas.

The actual EL training and assessment of library staff members is not a well-documented area in the professional literature. This severely restricts the ability to quantifiably evaluate or measure the scope of the training. No peer-reviewed studies document what EL baseline concept level expectations public library administrators hold for library professionals or what levels are needed to maximize the impact of EL skills development in children. Fasick and Holt (2008) in their book, *Managing Children Services in the Public Library*, offer credence to an assumption that ongoing staff training is not consistent or standardized. They write “some library administrators make all decisions about how to deliver staff training, others leave the decision to the discretion of the department head” (Fasick & Holt, 2008, p. 48). Fasick and Holt posit “although testing is rarely an acceptable way to check what a staff has learned, a survey evaluating training sessions might be helpful in assessing how well the training program works” (2008, p. 49).

Providing examples of library staff training initiatives, Stephanie Schauk, with the Maryland State Library, described emergent literacy training as being “tied to the
PLA initiative and linked to Maryland education standards” (2004, ¶). A peer coaching wiki, established in 2008 by Maryland librarians, is an example of an additional training resource that allows practitioners an opportunity to communicate and share emergent literacy ideas with colleagues (Emergent Literacy Peer Coaching Wiki, 2009). If it is assumed that the parents of the young children are in need of being taught emergent literacy concepts and activities, then it may also stand to reason that there should be an expectation that the librarians conducting the training sessions understand emergent literacy concepts. This is not to say that librarians conducting early literacy programs do not have high early literacy concept understanding, but rather, that there is a lack of well-documented attempts to access these levels using research-based tools.

Libraries have also turned to consultants for staff training programs. Mother Goose on the Loose (MGOL), as mentioned above, is an early literacy program designed primarily for librarians, early childhood teachers, daycare providers, and parents (Diamant-Cohen, 2005). According to the MGOL website, the program description states (2005):

Eighty percent of the material is repeated from week to week. This makes it easy for the leader to run programs on a weekly basis, since there is no need to recreate totally new programs each week, or to find new props to use each time. Brain research shows that repetition is one of the best ways for children to learn; children will get greater benefits and have more fun by returning to material that they have already heard than by being introduced to new materials each week. (¶)

This program was included as an example of early literacy programs interpreting current research theory into application in the field. Using outsourcing for training purposes increases the need to investigate the alignment between early literacy library programs and the emergent literacy concepts as identified by emergent literacy research experts. Do practitioners understand research-based emergent literacy concepts?

Outlining the processes involved in creating public library reading programs for preschoolers, librarians Cabré i Ametllé and Müller-Jerina write, “Implementing an emergent literacy program requires the work of a professional librarian or someone similarly qualified as well as a support staff, if required” (2004, p. 14). While these
authors do not provide a definition of a *professional librarian*, this term is frequently associated with a terminal degree or a specialist’s certification that is granted through the American Library Association (ALA) accredited Masters in Library and Information Science (MLIS) program.

Coursework and degree requirements for an MLIS degree are not consistent across accredited programs. It is quite possible for an accredited public-library children’s librarian not to have had a single course related to child development or literacy theory. Martinez (2005) discusses the lack of library course preparation related to a child’s developmental process of reading and writes, “Maryland’s public librarians attend a professional development session…to learn what state education leaders expect of children entering public kindergartens” (Martinez, 2005, p. 11). Professional degree program certifications and state requirements affect the level of child development and literacy-training modules a school media specialist must complete, but children’s librarians in public libraries do not have the same certification structure.

To date, a comprehensive assessment of how emergent literacy topics are integrated into the current accredited library degree programs has not been published. While it may be difficult to assess course offerings that include a component of the emergent literacy concepts, there are examples of courses completely dedicated to emergent literacy. The development of MLIS courses serves as another indication of the importance being placed on early literacy connections with libraries. A few of the course offerings are included to demonstrate a trend to create MLIS courses is underway. All MLIS programs are not currently including EL course offerings. This variability points back to the unknowns related to understanding what practitioners know and do not know about early literacy concepts and even calls into question how the concept of early literacy is actually being defined.

The Kent State Office of Continuing and Distance Education offered a graduate one-hour credit course. They (Kent State Office of Continuing and Distance Education, 2009) list the Library and Information course description as:

ABC’s of Emergent Literacy: The Building Blocks for Infusing Early Literacy Strategies Into Preschool Storytimes: If you’re a youth services library practitioner, attend this workshop to learn the basics of emergent literacy and the
role it plays in preschooler’s literary development. You’ll discuss basic skills in emergent literacy development, identify children’s literature that demonstrates these skills and incorporate specific strategies within preschool story times. Ultimately, you’ll create a six-week series of programs that infuse emergent literacy strategies. (¶)

Another MLIS course example is offered through the School of Library and Information Studies at the University of Alberta, Canada. This one-hour credit course, LIS 598: emergent literacy has two prerequisites: LIS 501 (Foundation of Library and Information Studies) and LIS 502 (Organization of Knowledge and Information). The course description, according to instructor Taylor-McBryde's (2009) syllabus is:

This course will provide an introduction to early or emergent literacy research and how it can be used to enhance services for children from birth to age five, their parents and care-givers. Research will focus on language development and developmentally appropriate programming. The course will examine the critical role of parents and the supportive role of family, care-givers and librarians in developing early literacy (or pre-reading) skills. (¶)

The relationship between the prerequisites is not stated on the website; however, the instructor Taylor-McBryde (2009) includes the student objectives. The students will:

(1) Have an awareness of the wealth of research and resources available on children’s language development and emergent (or early) literacy;
(2) Be aware of current trends in emergent literacy or family literacy programming and be familiar with the variety of services for children and parents that public libraries can and do provide to augment early learning;
(3) Understand the critical importance and role of the parent and care-givers in early learning; and
(4) Be able to apply the research to developmentally appropriate programming through the acquisition of new rhymes, songs, and stories. (¶)

With growing popularity among accredited higher-education programs, it is not uncommon on Library and Information institutional websites to find a faculty member listing EL as a research interest. An expansion of early literacy MLIS courses supports a
continued perceived level of importance placed on the EL domain. However, these listings lead to more questions. What levels of expertise do they bring to the profession? How much expertise is required to facilitate course development and librarian education in this domain? How is this term operationalized related to other domains; e.g., Psychology and Education?

An additional perceived importance indicator of EL library connections is found in the publishing world. Some of the books published to assist librarians in developing library services and in managing children’s services in public libraries have expanded traditional topics to include emergent or early literacy concepts; e.g., *Early Literacy Storytime @ your library: Partnering with Caregivers for Success* (Ghoting & Martin-Diaz, 2006), *Babies in the Library!* (Marino, 2007), *Building Blocks: Building a Parent-Child Literacy Program at Your Library* (Snow, 2007). There is high variability in emergent literacy presentations in these available resources and some of the library service how-to-books have minimal direct connections to researched-based emergent literacy concepts. An author's familiarity with current evidence-based practices may be influential in a reader’s translation of theory to practice. Again, these conditions spur additional questions. How much understanding of the emergent literacy core concepts is needed for a librarian to be successful in assisting parents in developing appropriate emergent literacy skills in the home? What misconceptions associated with supporting early literacy are held by librarian practitioners? Do some librarians see emergent literacy as simply a new label for what has always been offered in young children story times? According to Higgins (2007)

The concept of emergent literacy is now used to describe literacy processes that occur in the family circle and during traditional storytelling presentations in the public library... Children learn how to attend to language by interacting and modeling others in the environment. For example, if a child ‘models’ holding a book and babbling as if reading, this behavior is considered emergent literacy. If a child repeats rhymes and finger plays conducted in story time, this is also considered emergent literacy. Although emergent literacy is the term for such early literacy activity, the idea is not new – story times in public libraries have existed for over 100 years. (p. 58)
Now, even more questions present themselves. How recent do the publications need to be to address best practice? Has citing expired or unsubstantiated early literacy beliefs compounded the difficulty librarians experience in applying best-practice research? For example, in a 2001 presentation to the International Federation of Library Associations and Institutions (IFLA) Fiore referenced Anderson, Hiebert, Scott, and Wilkinson’s 1984 Report of the Commission on Reading stating “research has also found that the single most important activity for building the knowledge required for eventual success in reading is reading aloud to children” (2001, p. 4). Randomized studies on reading aloud have yet to demonstrate that reading allowed is the single most important activity for long-term conventional literacy gains. It is known that phonological awareness and alphabetic knowledge have significant impacts on long-term conventional literacy gains in children. Fiore continues to state “libraries have responded by designing programs that encourage reading aloud and other such literacy activities” (2001, p. 4).

Fiore, a highly regarded library literacy advocate, is the author of the classic book *Fiore’s Summer Library Reading Program Handbook* (Fiore, 2005). This perhaps sensitive example was not included to discredit the service of Fiore to the library field. This illustration highlights factors that may compound or aggravate the ability of a practitioner to discriminate evidence-based research findings related to emergent literacy concepts when the term *research* has varying degrees of rigor. A possible disconnect may exist between the evidence-based research and application to practice.

Thus far, examples have been provided to highlight the perceived importance of EL connections to libraries through the commitment of resources to (1) children's services that include enriched EL programming, (2) partnerships (3) EL training programs, (4) MLIS course offerings, and (5) availability of publications with dedicated library and EL components. With these examples, the coexistence of an unknown variability has been highlighted to support the need for research in these areas and the importance of this study.

Following a money trail may be another way to demonstrate that EL library connections are perceived to be important. There are many reasons grants may be awarded to various organizations. One of the decisions to fund projects is related to the demonstration of a need. Since grants are being awarded to libraries for early literacy
services, it may be assumed that this also demonstrates a level of importance attached by the funding agencies to the stated goals of grant proposals. Since all public library children's EL programs are not being supported through outside grants it can also be assumed that there are different costs attached to the various levels of service offerings. Conducting early literacy library services may be an expensive endeavor. Services may extend beyond storytimes to include the purchase of EL materials, e.g., early literacy kits for patron circulation. Some libraries offer parent EL workshops and daycare EL workshops. Without a centralized accountability system in place, collecting accurate expenditures allocated to early literacy programs in conjunction with libraries is an extremely difficult, if not an impossible, task under the current reporting methods.

This next section illustrates some of the funded EL library connected projects. These are used to express an understanding that funding EL projects supports the perspective that EL connections are valued to have a perceived importance. Based on the available information regarding the allocation of funds it is not known what it costs to implement an effective EL storytime. Investments in EL programs do not appear to be administered consistently across all libraries; however, early literacy is often collapsed into the larger category of literacy. Funds allocated to facility space to house these various programs or to hire additional personnel may also not be reflected under current financial reporting systems. Funding disparities might spawn additional questions, e.g., how consistently are early literacy programs being applied across the United States? What about the communities with library staffs that do not possess grant writing skills? What effect does funding differences have on the EL training learning outcomes?

Funding, especially in a depressed economy, is competitive and limited. This funding disparity for libraries seeking to fund EL programs may translate into at-risk communities with literacy needs that are not adequately being met because of a lack of funding. Writing on behalf of the Central Texas Library System, Nyfelder reports that after attending the 2006 Texas State Library and Archives Commission sponsored public library emergent literacy training workshops, based on the Every Child Ready to Read @your library model, they “saw the need for an organized way to involve public libraries in teaching the importance of early literacy skills. We applied three times for grants to support library staff training and reaching parents/caregivers through community
partnerships” (Nyfeler, 2009, ¶). After failing to receive funding on multiple past proposal attempts, Nyfeler reported that the program received funding in August 2008 on the third try.

Funding for library emergent literacy initiatives is derived from private and federal sources. The National Home Library Foundation (NHLF), a small private organization for over 40 years, awards grants between $500 and $3000 (National Home Library Foundation, n.d.). In recent years, the NHLF Trustees have supported numerous funding requests for the purchase of children's books that include infant and preschool programs. For example, the DC Public Library Foundation was awarded $2000 “for the purchase of books for their emergent literacy program books kits, which provide resources for over 120 in-home child-care providers, reaching over 360 impoverished children” and $500 “to fund take-home books which supplement the emergent literacy kits program …” (National Home Library Foundation, n.d., ¶). This grant example points to some of the extend costs associated with early literacy programs, e.g., book kits, and demonstrates that library connections to early literacy extend beyond the physical library facilities. As librarians continue to expand the role of EL trainers beyond librarians, e.g., training caregivers and child care providers, it becomes increasingly important to access the core knowledge levels of the trainers. At this time, it is not known what is known.

The Institute of Museum and Library Services (IMLS) funds research and programs that promote emergent literacy and library initiatives through the Library Services and Technology Act. The IMLS allocates money to each state based on an established formula. The states accept proposals and distribute the money. There are numerous examples of state funds awarded to public libraries with proposals that include objectives that emphasize pre-literacy for pre-school children and their families. State Libraries report back to the IMLS on how the money was allocated; however, as of September 2009, the current tracking reporting measures did not designate early literacy or emergent literacy initiative categories.

The demonstrated emergent literacy funding connections with library services highlights the importance of asking a few basic questions: (1) How knowledgeable are the people conducting the emergent literacy training? (2) How much learning takes place during EL training? (3) How much of this training translates into practice? and, (4) To
what extent are the early literacy concepts research-based?. This study does not address these questions directly; however, the availability of the research-based instrument from this study will support future initiatives designed to investigate these types of issues.

1.2 Importance of the Study

Current reading research literature supports exposure to emergent literacy concepts as an important factor in a young child’s development of successful lifelong literacy skills. It is known that public libraries have placed a perceived level of importance to EL library connections. Public library connections have been shown to include the (1) development of early literacy programs, (2) creation of partnerships with early learning organizations, and the (3) allocation of resources for early literacy facilities, personnel, professional development, and materials. Public libraries are also known to be conducting EL outreach and training for schools, families, and child care providers. Public library researchers and public librarian administrators have sought and received grant money to support these EL library connection initiatives.

There are some known indicators that point to a perceived importance in children’s programming services. For example, (1) children program attendance at public libraries increased by 13.9% between 1999 and 2008 (Institute of Museum and Library Services, 2010), (2) 552 of the 889 library systems reported having provided literacy programs for children ages birth to five (Public Library Data Services, 2010), and (3) the appearance of early literacy as a stated IMLS stated 5-year plan mission and goal statements tripled between the 2003-2007 and 2008-2012 reports. For an expanded discussion on early learning and literacy in the National landscape see Dresang, Burnett, Capps, and Feldman (2010).

Public libraries are talking about early literacy. There are known pointers that indicate a perceived importance; however, there are many more unknowns that raise the need to conduct research in the area of EL and public library connections. What does early literacy look like in public libraries? What impact do these programs have on long-term conventional literacy gains? There is a need to conduct evidenced-based research literature to investigate EL library connections.
In much the same way that the emergent literacy reading research community benefited from the National Early Literacy Panel task force, the library professional community needs to take inventory of the emergent literacy, core knowledge levels of adult practitioners working in an environment with expectations of supporting literacy development of young children. The EL-Capstone will provide a research-based instrument that will assist the field in establishing baselines.

In turn, the information gathered through future applications of this tool may be beneficial to organizations seeking future funding opportunities. Understanding the EL core levels may also affect accredited higher education curriculum reform movements in library and information programs. Ultimately, knowing emergent literacy practitioner baselines may support initiatives designed to expand and/or to affect literacy opportunities for young children.

This instrument is important to the library profession in collecting a baseline of existing emergent literacy capstone levels. Before courses become a standard in the accredited library education programs, measuring what is known and what is not known becomes important. As President Obama’s administration evaluates existing programs and creates new funding sources, librarian professionals demonstrating appropriate qualifications may be able to improve their funding opportunities and, as such, may allow public libraries to better serve needs of the communities.

A decade of fragmented library and emergent literacy initiatives is indicative of the value behind the importance of developing an instrument to inventory the knowledge levels of practitioners working in early childhood environment. Some public librarians have attended an American Library Association (ALA) accredited program to earn professional degrees in Library Science. Again, more questions are created. How prepared are certified public librarians to work in early childhood literacy environments? Are public library emergent literacy programs built on current research-based outcomes? Are grants and other funding sources being used to duplicate emergent literacy outreach programs? These questions, and others like them, that fuel future research opportunities, will benefit from the development of the EL-Capstone. The development of this research-based-instrument will provide a tool to assist the field in identifying where librarians are as a profession related to emergent literacy core knowledge.
Awareness of emergent literacy core knowledge levels across the librarian profession holds the potential to positively impact librarian contributions to the early literacy skills development within a community. Understanding how prepared librarians are in being able to serve an emergent literacy leadership role in the community’s literacy initiatives may inform learning outcome goals in accredited library degree programs and professional development training. An instrument to measure these emergent literacy levels needs to be available to the library professions. Before research is implemented to measure learning outcomes of young children attending early literacy programs associated with libraries, it is important to first assess the EL capstone levels of the adults delivering the information to the children. Therefore, the goal of this dissertation to develop an instrument to measure emergent literacy, core knowledge concepts of librarian practitioners is an important endeavor and it contributes to need to research EL library connections. This study directly enhances Library and Information education research by providing a valid and reliable instrument for assessing the emergent literacy, concept knowledge levels.

1.3 Research Goals

Overall, the goal of this research was to design a valid and reliable, community-sensitive tool to support the need of funded agencies to assess emergent literacy understanding in adults that may be called upon as leaders to promote emergent literacy development of young children. Secondary goals were to answer the following research questions:

1. What general concepts are identified as emergent literacy concepts during interviews with library practitioners associated with public libraries?
2. What emergent literacy topics and subtopics are identified through test theory as discriminating emergent literacy core knowledge concept test items for adults working within early childhood environments?

1.4 Limitations

The primary objective of this work is the development of a research-based instrument (EL-Capstone) that may be used in future research investigations designed to
measure the emergent literacy, core knowledge levels of adult practitioners working in an environment with expectations of supporting literacy development of young children. The EL-Capstone is not a predictive tool to determine the effectiveness of a high score verses a low score performer. Future research may be supported by this concept inventory to gather information that may be used to inform library degree programs, library institution administration, policy makers, and funding institutions to current emergent literacy concepts capstone levels of practitioners and students. Beyond informing the design of the instrument, participants' individual scores are not generalizable to larger populations.

The participant feedback data collected through the cognitive interviews with librarian practitioners are not generalizable to other librarian populations. The small sample of convenience was only used to provide information related to the readability of the instrument; however, this selection process may have affected the nature of the results obtained during the interviews. It should be noted, the individual test scores are not interpreted related to the person’s learning experiences and this data did not attempt to explain the level of emergent literacy concept understanding of other populations of librarians or practitioners. In addition, steps were taken to promote honesty in responses but the data are still subject to participant motivation and response bias.

All of the participants involved in this study are volunteers and this could have data implications. Again, the participants' responses and scores in all phases of EL-Capstone’s development are not generalizable to other populations. Future research designs are needed to extend the validity of this instrument.
CHAPTER 2

LITERATURE REVIEW

As described in the introduction, this literature review introduces and defines characteristics of assessment instruments identified by the term *concept inventory* (CI). Past approaches to CI instrument construction reveal variability in instrument designs and in the measurements chosen to demonstrate validity and reliability.

Commonly acknowledged as the catalyst for subsequent CI research, the Force Concept Inventory (FCI) has been presented in detail before highlighting some of the additional CI instruments. The FCI was frequently referenced in the concept inventory literature and it served as a baseline instrument when describing methodological deviations. The additional inventories included in this literature review had unique design characteristics that had the potential to be influential during the EL-Capstone construction or analysis phase.

The scope of this dissertation was limited to the initial development of a valid and reliable instrument. Two test theories provided the theoretical framework of the EL-Capstone: classic test theory (CTT) and item response theory (IRT). While these two theories informed the decisions related to format and application of statistical measures needed during the design and analysis phases, many of the CTT and IRT measures were not required during the initial development phase of EL-Capstone because a unique population was not being assessed in this particular study.

2.1 Concept Inventories

Measuring attitudes and information levels of a person or group of people appears in the literature under numerous terms. For example, these instruments are called scales, indexes, measures, questionnaires, batteries, or inventories. Sometimes, they are known simply as a test (i.e., Nutrition Knowledge Test). Standardized tests or common assessments are typically designed to rank the test-taker’s content knowledge or to measure changes in learning outcomes.
The primary scope of this review covers assessment instruments identified by the term concept inventory (CI). Appearing similar to a common assessment in format, a specificity classification assignment of concept inventory may occur when the primary goal is to collect information to inform educators and policy makers of domain specific capstone levels.

Similar assessment instruments directly related to emergent literacy that were not classified in the general research literature as a concept inventory were also addressed within this section. Even within this narrowed scope, the literature on past approaches to CI instrument construction revealed variability in both the instrument design and the measurements chosen to demonstrate validity and reliability.

The use of concept inventory diagnostic test items to assess conceptual understanding has been successful in the sciences and engineering domains (Allen, 2006; Bailey, 2006; Bowling et al., 2008; Christian, Belloni, & iLumina, 2001; Garvin-Doxas, Klymkowsky, & Elrod, 2007; Hestenes, 1992; Knudson, 2006; Lindell & Sommer, 2004; Mulford & Robinson, 2002; Richardson, Morgan, & Evans, 2001; Smith, Wood, & Knight, 2008; Stone, 2006). Similar instruments have been developed in other domains as well, e.g., mathematics (Almstrum, 1999; Epstein, 2006) and teacher preparation (Almstrum, 1999; Flowers & Algozzine, 2000). Concept inventories afford an opportunity to standardize the assessment of conceptual knowledge through the use of common validated and reliable instruments.

The success of concept inventories is not consistent across domains. Garvin-Doxas, Klymkowsky, and Elrod posit that the acceptance of an instrument by field leaders and particular domain cultures may be influential in the support of a particular concept inventory instrument (Garvin-Doxas, et al., 2007). These researchers, for example, speculate that the physics inventory acceptance may be due to the inherent nature of physicists and in their acceptance of key introductory postulates. Biologist on the other hand, may find agreement on just the foundational elements alone to be a challenge (D'Avanzo, 2008; Garvin-Doxas, et al., 2007; Michael, 2007). This difficulty in agreement has not deterred the development of validated biology inventories; e.g., the Conceptual Inventory of Natural Selection (Anderson, Fisher, & Norman, 2002) and the Biology Concept Inventory (Garvin-Doxas, et al., 2007).
Some concept inventories are being used for a pre-test and a post-test to compare changes in student learning while other concept inventories are more diagnostic and investigate changes in students’ preconceived notions related to the major concepts (Klymkowsky & Garvin-Doxas, 2008). These preconceived misconceptions may be exemplars held prior to any formal instruction. For additional discussion on concept inventories and the persistence of misconceptions related to learning outcomes, see Morton, Doran, and MacLaren's (2008) article, "Common Student Misconceptions in Exercise Physiology and Biochemistry" or Almstrum's (1999) article, "The Propositional Logic Test as a Diagnostic Tool for Misconceptions about Logical Operations."

Concept inventories, as pretest/posttest measures, are assisting researchers within learning environments to compare the learning outcome differences between people arriving with already preconceived, faulty, or incomplete conclusions about a topic compared to people that are new to a concept. Many of these instruments are designed for diagnosis and for evaluating learning outcomes (Anderson, et al., 2002; Corkins, 2009; Libarkin, 2008). Concept inventory instruments are available in multiple phases of development. It is suggested that prior to the administration of any psychometric measure, the current status of the instrument is validated for assessment outside the original sample populations.

A frequently used CI format is a multiple-choice instrument. Usually embedded in this format is the inclusion of common misconceptions in the item response choices. Concept inventories tend to be narrow in scope and numerous concept inventories may exist even within a single domain; e.g., physics (Ding, Chabay, Sherwood, & Beichner, 2006; Hestenes, 1992; Thornton & Sokoloff, 1998). There are exceptions to the narrow focus, e.g., the Conceptual Survey of Electricity and Magnetism (Maloney, O’Kuma, Hieggelke, & Van Heuvelen, 2001) and the Chemistry Concept Inventory (Krause, Birk, Bauer, Jenkins, & Pavelich, 2004).

The idea of creating an instrument to inform pedagogy is built into the framework of several concept inventories (Hestenes, Wells, & Swackhamer, 1992; Libarkin & Anderson, 2007). Literature searches conducted between January 2008 and September 2009 did not retrieve concept inventories that have been published within the area of emergent literacy. As such, an overview of concept inventories with examples of specific
instruments is included in this work since the information is foundational to the instrument design decisions for the EL-Capstone.

2.1.1 The Catalyst: Force Concept Inventory (FCI)

Unsatisfactory student outcomes from introductory physics courses inspired Hestenes (1987) to investigate the inefficiencies of traditional physics instruction. Referencing the philosophy that lead to the design of the Force Concept Inventory (FCI), Hestenes posits, “The question is not whether students can learn physics, but whether instruction can be designed to help them learn it more efficiently” (1987, p. 440). The 1992, physics diagnostic instrument, FCI, evolved over time from the earlier work by Halloun and Hestenes (1985) and the initial results have been repeatedly replicated. The FCI is generally credited as a catalyst in the success of concept inventories (Allen, 2006; Libarkin et al., 2005). For this reason, the FCI serves as the demarcation point for purpose of this literature review.

The 30-item multiple-choice questions in the FCI are organized into six categories: Kinematics, First Law, Second Law, Third Law, Superposition Principle, and Kinds of Force. Each category has approximately five questions. Using a multiple choice format, each question has one correct response and three incorrect responses. The incorrect responses are assigned to misconception sub-topics: Kinematics, Impetus, Active Force, Action/Reaction Pairs, Concatenation of Influences, and Other Influences on Motion. Figure 1 highlights the importance the FCI places on collecting information concerning student misinformation. Informing educational practice is a central theme for the FCI.

The ability to parse Newtonian and non-Newtonian concept acquisitions are embedded in the misconception sub topics. The contextual nature inherent with the use of misconception distracters drew validity concerns of early data interpretations as the factor analysis used only higher scoring student data (Heller & Huffman, 1995; Rebello & Zollman, 2004). In the administration of FCI at Arizona State University, the student’s final course grade in introductory physics was highly correlated with post test scores ($r=0.56$, $p=0.0001$). This predictive measure was not highly correlated when FCI was administered at a different university; however, the instrument was not validated.
previously in the additional institution and controls were not in place to capture teaching methodologies (Hestenes, et al., 1992). Instructional pedagogical influences are incorporated into the goals of the FCI. The FCI typically reports to have internal consistency Cronbach’s alpha coefficient values of 0.86 on the pretest and posttest values of 0.89 (Halloun & Hestenes, 1985).

![Diagram](image)

**Figure 1.** A basic schematic of the Force Inventory Concept illustrating the six major categories and the central role of student misconception sub-topics in the response options.

The FCI has been successful as a measure when comparing traditional teaching and interactive engagement (Hake, 1997). Michael credits the FCI with an ability to demonstrate “…that students capable of producing correct solutions to complex quantitative problems nevertheless seem to have a poor understanding of the concept underlying equations they are using” (2007, p. 389). Even though the FCI elicits information about misconceptions in mechanics, Hestenes cautions against seeing this as the primary function of the instrument that is based on a conceptual model (2006).
2.1.2 Influential Exemplars

The success and utility of the FCI spawned additional research investigations into future applications for concept inventories. The inclusion of specific concept inventories in this literature review is not an indication of the popularity or success of any particular inventory. The selected instruments are representative of concept inventories and they are being used to illustrate the variability of concept instrument designs and applications. The instruments are listed alphabetically to strengthen the ability of this literature review to serve as a reference tool during the development of the EL-Capstone instrument. Duplications of any concept inventory acronyms, included in this review, will appear with numerical subscripts to aid readability; e.g., the Statics inventory and the Statistics inventory are both known as SCI; therefore, they will be re-assigned new acronyms of SCI\textsubscript{1} and SCI\textsubscript{2}.

2.1.2.01 Basic Technology Competency for Educators Inventory (BTCEI)

As technology moved into the classrooms, teacher education programs developed a need to evaluate the curriculum. To inform the development process of teacher preparedness for technology in the classroom integration courses, Flowers and Algozzine (2000) developed the Basic Technology Competency for Educators Inventory (BTCEI). These researchers do not use the term concept inventory explicitly in the name of their instrument; however, the goals and objectives of the BTCEI are aligned with those of a concept inventory and therefore, it warrants inclusion in this literature review as an example based outside the science domain.

This instrument affords educators an opportunity to have a valid instrument to assess teacher education students’ basic technology skill competencies. They developed the items through “a consultation of the literature and a review of the fundamental concepts and skills established by the Association for Educational Communications and Technology and the International Society for Technology in Education” (Flowers & Algozzine, 2000, p. 412). During the scale development process, faculty experts reduced the 207 potential instrument items down to the 36 items that are included in the BTCEI. While Flowers and Algozzine use faculty as experts during the construction of BTCEI,
the selection process and the number faculty experts that participated in the study are not noted in their 2000 publication (Flowers & Algozzine, 2000).

Students enrolled in a required university introductory education course were purposefully partitioned into three groups for participation in one of the three instrument development phases: (a) concurrent validity study (n=107), (b) experimental study (n=54), and (c) test-retest (n=116). During the concurrent validity study, the participants completed the BTCEI and the results were compared to a hands-on, performance-based test in which the participants were expected to demonstrate basic technology skills.

The BTCEI was administered before and after course instruction; therefore, the test-retest phase administered to 54 participants may more accurately be characterized as a pretest/posttest that used two posttest administrations within a 2-week span. The multiple posttest phase was included to measure the score consistency of the instrument. The BTCEI is a 4-point competency scale that is accompanied by a brief scale definition to improve consistency of interpretation of item response options by the participants. It does not attempt to capture common misconceptions through the use of a multiple choice, distracter format.

This instrument pilot test indicated that clustering the items by subcategories, rather than random presentation of items, reduced the participant inventory completion times. The researchers, Flowers and Algozzine, noted that factors such as effects of fatigue and practice effects were secondary to the benefits of conceptual clustering (Flowers & Algozzine, 2000).

2.1.2.02 Chemistry Concept Inventory (CCI)

Using a concept inventory to help understand students’ reasoning approaches to solving complex problems is foundational to the work by Krause, Birk, Bauer, Jenkins, and Pavelich (2004). The topics covered in the Chemistry Concept Inventory (CCI) were developed through a collaborative effort between chemistry educators and engineering faculty. The questions and misconceptions response options were selected through an extensive literature search of previous chemistry education research. These were all conceptual in nature and the test items did not depend on mathematical or previously memorized facts.
With at least three questions per sub-topic, the original multiple-choice test with 30 questions (Version A) was reduced to 20 questions (Version B) to facilitate a shorter administration time. The questions determined by the researchers to have the least negative effect on the coefficient alpha (Cronbach's $\alpha$) were eliminated and they determined that by using two questions per sub-topic, the reliability coefficient inherently is reduced (Krause, et al., 2004).

At the test item level, the CCI internal reliability was measured using difficult indices and discrimination values. The alpha values were calculated to insure that the results were significant for the whole test and that the outcomes were not simply due to chance. Krause et al. noted that their inventories were reasonably reliable with scores above 0.7 in an alpha value range of 0 to 1 (Krause, et al., 2004).

To help establish validity, the researchers recruited first year chemistry students (n=11) to participate in interviews specifically designed to expose misconceptions, problem-solving approaches, and instrument readability. On an iterative basis, the instrument questions were slightly modified (Version C) and then administered to an undisclosed number of graduate teaching assistants before being released as a freely available concept inventory.

It should be noted that the research questions behind the development of the CCI included investigating a correlation between course performance and CCI scores. Version A was used for this data and the students grades were directly connected to the CCI scores since the instrument was used as a graded quiz (Krause, et al., 2004). Other instrument researchers caution that a concept inventory is not a test of ability but an assessment of a much broader conceptual understanding of a specific. As such, connecting concept inventories directly to student course grades may introduce lurking variables into the analysis.

2.1.2.03 Conceptual Survey of Electricity and Magnetism (CSEM)

The Conceptual Survey of Electricity and Magnetism (CSEM) is a broad, 32-question, multiple-choice, survey instrument that takes a twist in defining the term concept. The scope of the instrument's concepts crosses into multiple domains, i.e., force, motion, and energy. Maloney, et al. interject a perspective that “in the domain of
electricity and magnetism, most students lack of familiarity with both the phenomena and most of the concepts, language, principles, and relations” (2001, p. S12). Therefore, the stated goal for CSEM is to acquire an overview of students’ knowledge and to capture a student’s schema or a preconceived notion of topics in electricity and magnetism (Maloney, et al., 2001). In this concept inventory instrument, the researchers developed the CSEM tool to capture the phenomena rather than the formalism or specific concept instruction.

The design of the instrument began with experts (defined as a group of experienced college physics professors). The preliminary instruments tested for one academic year began as two independent concept inventories – one for electronics and one for magnetism. The inclusion of open-ended questions in the early versions of CSEM allowed for data collection of student misconceptions and provided the researchers with common usage vocabulary that reflected student terms of the topic.

The instrument analysis varied with the progression of instrument. For example, with pretest scores relatively close to random guessing, Maloney et al. placed value only on the difficulty and discrimination measures of the post-test scores of the CSEM version G (2001). Validity estimates for CSEM included the collection of appropriateness and reasonableness rankings by physics faculty from two-year colleges (n=42).

CSEM's reliability included a principal components factor analysis and the use of KR20. Although Maloney et al. found the KR20 (Cronbach's alpha) calculation to underestimate reliability, they state that the "formula gives a representation of the average correlation between the test subdivided into two halves in all possible ways" (2001, p. S14). Their posttest estimates are approximately 0.75 (a strong value since an acceptable group estimate score range is 0.5 to 0.6). Version H of the Conceptual Survey in Electricity and Magnetism is included as an appendix in the Maloney et al.'s (2001) article, "Surveying Students' Conceptual Knowledge of Electricity and Magnetism."

2.1.2.04 Genetics Concept Inventories

Several independent genetics concept inventories are available as an example of variations in instrument construction within the same domain that still lead to reliable and valid instruments. The Genetics Concept Inventory (GCI) developed by Hott has goals of
being a curriculum assessment tool for measuring student concepts and for informing policy decisions (2006, p. 20). This instrument is included as an appendix in Hott's 2006 dissertation.

This instrument was designed to measure six genetics concept areas. Each area has a minimum of two test items to strengthen the instrument's ability to measure the concepts. The GCI was designed for use with introductory biology undergraduate students. Because of the inherently diverse nature of this group, Hott conducted two focus groups to check for potential vocabulary or wording difficulties.

The original GCI, comprised of 25 content and eight demographic questions, was administered as an online pretest and posttest. Decisions to remove outlier participants from pretest and posttests were made for extremes in test completion times. The remaining participants had a pretest mean of 11 minutes to complete the 33 test items. The posttest results for completion time were 11.5 minutes. The questions on the posttest were presented in an alternative order to help control for test/retest effects.

Hott's data analysis of the content posttest items included several reliability measures. An acceptable Cronbach's alpha estimate was set for 0.06 and the standard classical test theory (CTT) discrimination and difficulty measures were completed. Differences in pre-existing skill level influences were captured through an overall scores comparison between upper-level biology majors and non-science major participants.

Hott tested for gender bias by using the Mantel-Haenszel statistics analysis. A one-way ANOVA provided data results for possible differences in the participants' confidence and background in science plus academic standing. Hott also collected the amount of textbook coverage and the amount of instructor time spent on each of the six content areas for comparison with the participant scores. The overall pretest/posttest comparisons were completed for 130 matched participants.

A similar instrument developed to measure genetics concepts is the 25-question Genetics Concept Assessment (GCA). This instrument's developmental processes included a review by experts, participant interviews, and a pilot study; however, the primary stated objective is to measure student learning goals with the pretest/posttest instrument (Smith, et al., 2008).
An alternative genetics concept inventory is the Genetics Literacy Concept Inventory (Bowling, 2008) which is more commonly called the Genetics Literacy Assessment Instrument (GLAI). The acronym GLAI will be used in this review. Advances in genetics and biology have impacted the need for the general public to be able to make informed decisions related to these concepts. The GLAI is designed to provide a tool that will help "access students’ understanding of genetics concepts and their level of genetics literacy (e.g., genetics knowledge as it relates to, and affects, their lives)” (Bowling, et al., 2008, p. 15). The Concept Inventory of Natural Selection (Anderson, et al., 2002) informed the GLAI research design. The GLAI is included in this literature review to illustrate that while there are commonalities in concept inventory research designs, an instrument's intended use will heavily influences the execution of these process steps.

In describing the GLAI instrument design, Bowling et al., write, “The steps of the process involved defining the content, development, and selection of test items, review by professionals, focus group interviews, pilot study data collection, and evaluation” (2008, p. 16). The starting point for defining the content began with an American Society of Human Genetics published list of genetics benchmarks that contained six main concept areas and 43 sub-concepts. The authors reduced this list to reflect only areas that were sensitive to the literacy goals of the instrument. While several biology textbook test bank resources along with some commonly used biology exam test questions were referenced during the early stages of development, the actual questions used in GLAI underwent numerous revisions. Examples of the question revision transformations and final GLAI test items are available in the Genetics Education publication, "Development and Evaluation of a Genetics Literacy Assessment Instrument for Undergraduates" (Bowling, et al., 2008).

The test-item pool of 56 questions was divided into two feedback forms and distributed for review. Twenty-five of the 73 invited genetics professionals, instructors, and graduate students agreed to participate in the study. These reviewers completed a 3-question per test-item feedback form that was converted into quantitative data and they provided qualitative data in the form of verbal feedback related to the wording and general question construction. Both the quantitative and qualitative data were analyzed
and used to inform the decisions for test-item revisions and for the selection of the 33 test items used in the initial version of GLAI (Bowling, et al., 2008).

Nine self-selected undergraduate students agreed to participate in one of the two focus groups. Via personal response systems, the focus group participants answered the GLAI test items while responding to a series of follow-up probing questions designed to expose any misinterpretations related to the conceptual intent or wording of the questions. The focus group findings informed revisions.

The post-focus group, revised GLAI was administered one time to 11 self-selected undergraduates from an introductory biology course. These results lead to more revisions. The 31 item, revised pilot test, version of GLAI was administered as a pre-course test (n=395) and as a post-course test (n=330) to undergraduate students enrolled in introductory biology and genetics courses. As a comparison, GLAI was administered as a pre and post test to introductory psychology students (n=113). For a baseline comparison to people with known high levels of genetics content knowledge, the GLAI was administered one time to genetics graduate students (n=23).

Elrod used the GCI, comprised of "38 questions: 36 multiple answer/choice, true false, fill-in-blank, matching and 2 short answer questions" (Elrod, 2007), to assist in the measurement of student genetics literacy outcomes based on different learning environments. The instrument was administered to undergraduate students through the use of Blackboard tools; however, Elrod's results were not analyzed with any non-Blackboard inherent statistical methods.

2.1.2.05 Geoscience Concept Inventory (GCI)

The Geoscience Concept Inventory (GCI) was a collaborative team design piloted on over 3,500 students from across diverse higher-education settings offering geoscience courses (Libarkin & Anderson, 2005). The statistical use of a single Rasch scale and differential item functioning to flag item bias separated GCI from earlier concept inventories. Highlighting the importance of the single Rasch approach, Libarkin and Anderson posit “all GCI sub-tests can be tied to a single scale, allowing faculty nationwide to have the freedom of designing tests relative to their course content, without sacrificing the ability to compare student learning and instructional approaches used in
different courses” (2007, p. 148). While several of the earlier concept inventories had test items that were based on predetermined content, the GCI coded student interviews and student questionnaires to inform a portion of the test items.

The GCI, administered to assess undergraduate student learning in entry-level geoscience courses, emphasized capturing conceptual changes (thus an assessment component) but GCI was also designed to diagnose misconceptions that students brought with them into the classroom (Libarkin, 2008; Libarkin & Anderson, 2005; Libarkin & Anderson, 2007).

In their findings, Libarkin and Anderson noted “in a study of 43 courses and 2500 students, we find that students are entering geoscience courses with alternative conceptions (sometimes called "misconceptions"), and in many cases are leaving the classroom with these alternative ideas intact” (2005, p. 394). Conducting interviews and administering open-ended questionnaires, Libarkin, Anderson, Dahl, Beilfuss, Boone, and Kurdziel had similar findings in an earlier qualitative analysis of college students’ ideas about the earth (2005).

2.1.2.06 Signals and Systems Concept Inventory (SSCI)

The Signals and Systems Concept Inventory (SSCI) developed in 2000 was initially designed to measure students understanding of fundamental concepts in linear systems with a primary objective of evaluating curricular reform and pedagogical techniques (Wage & Buck, 2001). Instrument test-item samples are included in the article, "Signals and Systems Concept Inventory" (Wage, Buck, Cameron, & Welch, 2005). The developers acknowledge the inspiration of the Force Concept Inventory in the development of SSCI.

The SSCI inventory consists of 25 multiple-choice questions with an emphasis on five concepts. The instrument's emphasis on measuring conceptual understanding, and not on the level of computational skills, did not remove design considerations for addressing wording issues related to variations in the use of mathematical notations or with known variances in common terminology. The authors provide an example of the terms radian frequency and hertz being used by different university textbooks when
discussing the identical concept (Wage, et al., 2005). By design, the SSCI test items rely heavily on figures and diagrams and tend to be less numerically-based.

To aid in the development of common misconceptions that may be used in the answer options for the SSCI test items, the early versions of this instrument allowed the participants to select none of the above as an answer and then write in a response. The collection of misconceptions from this method could facilitate the collection of natural language terms that may be used to inform the final wording of the test items. The developers of SSCI report that few alternative write-in responses were made; however, they suggest that this may be an indication that the already available misconception answer response options covered the common misconceptions (2005).

A discrete-time and continuous-time version of the SSCI is available. Wage et al, write that "striking outcome of the SSCI research study is the normalized gain analysis showing that students in traditional lecture courses master (on average) only 20% of the concepts they do not know prior to the start of the course" (Wage, et al., 2005, p. 460). The SSCI website (http://signals-and-systems.org/) is designed to promote and encourage research work associated with the use of this instrument.

2.1.2.07 Statics Concept Inventory (SCI)

Danielson and Mehtra developed foundational statics taxonomy with the initial support of eight additional mechanic engineering educators (2000). Danielson notes that a Delphi process was initiated to assist in the separation of potential test items that would be classified as concepts rather than as skills (Danielson, 2004). The Delphi study was an iterative process involving approximately 18 participants recruited from a mechanics list serve (Danielson, 2004). The first round of the Delphi was completed through e-mail communication. The participants agreed upon term definitions and they were able to separate the concepts from the skills. A dedicated website was set up for the second round in which the participants rated the importance of the concepts and skills for Statics students. Danielson points out the strength of a Delphi study since the participants were to complete the task and they were able to identify additional concepts that had been missed prior to initiation of the Delphi study (2004).
Steif and Dantzler (2005) expanded the work of Danielson and Mehta and created a new instrument, the Statics Concept Inventory (SCI). This multiple-choice instrument contains 27 items addressing a total of eight concepts and includes four misconception detractors in the answer response options. This instrument was tested with college undergraduates across five diverse university campuses. The test was administered at the beginning of the course and at the conclusion of the semester. A stated outcome for the SCI is to capture data that may help examine the relationship between undergraduate statics problem-solving abilities and conceptual understanding (Steif, 2003). The first version of SCI had 20 test items and the 81 undergraduates enrolled in the Statics course were given 35 minutes to complete the instrument.

Multiple testing environments have since been used for the administration of SCI. One of the research participant sites used a pencil and paper format with a 50 minute time constraint. The students (n=245) at the remaining four test campuses received a portable document file (pdf) to download, complete, and return electronically to the researchers. Participation credit received by the students was independent of the actual test-score outcomes. Steif and Dantzler note “it was impossible to monitor the time students spend taking the test, or whether they had help while doing the test” (2005, p. 6).

A difficulty index was not assigned prior to administering the inventory but it was captured during the item analysis through the percentage of students who answered the inventory item correctly. A discrimination index set to exceed 0.3 was achieved on all but one of the instrument items. The inventory's reliability measurements were determined to have a strong Cronbach's alpha coefficient of 0.89 (Steif & Dantzler, 2005).

Steif and Dantzler compared their students' SCI total scores to their final course grades to add criterion-relater validity that may help to predict student success in a Statics course. Standard statistics for tracking the pretest-posttest performance changes included the mean, standard deviation, maximum, minimum, and median values. Steif and Dantzler posit that "this test does appear to measure an ability which can change (markedly) with a semester studying Statics, affirming its [SCI] use as a tool to capture gain in conceptual understanding" (2005, p. 371)
2.1.2.08 Statistics Concept Inventory (SCI$_2$)

The intent of the pre-post test format of the Statistics Concept Inventory (SCI$_2$) is to capture student conceptual knowledge changes in introductory engineering statistics courses to compare at a national level. SCI$_2$’s emphasis shifts away from measuring learning outcomes in a traditional classroom assessment to assist in the accreditation criteria as defined by the Engineering Accreditation Commission (Allen, 2006). A broader outcome, according to Allen, is “to draw pedagogical implications leading to an enhanced understanding of statistics education” (2006, p. 11).

Many of the earlier concept inventories were designed with a narrow domain focus. The SCI$_2$ is an example of a concept inventory designed to cover a broader scope. The concepts are partitioned into five subtopics: probability, descriptive statistics, inferential statistics, graphic interpretation, and advanced topics. The SCI$_2$ is a multiple-choice format with distracters used to identify commonly held misconceptions.

A 2002 pilot study (n=139) of three instruments was administered to five groups across Statistics courses from multiple departments to measure the broad standardized assessment goals that are central to the SCI$_2$. The participants were given 35 minutes to complete the 32 test items, SCI$_2$ instrument. A Statistics attitude survey and a basic demographic questionnaire provided additional information that was used during the analysis to compare differences in the population sample, e.g., gender, race, discipline, mathematics background. Stone et al. calculated an internal reliability coefficient of 0.578 for the pilot study (2003). The results of the pilot study, focus groups, faculty surveys (n=23), and item analysis were used to inform the test-item revisions and to validate the instrument for content validity (Allen, 2006; Allen, Stone, Rhoads, & Murphy, 2004).

The researchers correlated SCI$_2$ scores with course grades obtained in Statistics courses from both the engineering department and mathematics department. Concurrent validity was originally attained with the engineering, statistics course, participants but not with the mathematics, statistics course, participants (Allen, et al., 2004; Stone, et al., 2003).

The Statistics Concept Inventory is well documented through presentations, publications, and two doctoral degrees granted by the University of Oklahoma. Allen's

2.1.2.09 Teaching Perspectives Inventory (TPI)

The Teaching Perspectives Inventory (TPI) was developed to quantitatively measure and identify teaching concept perspectives and to provide a baseline of information (Pratt, Collins, & Jarvis-Selinger, 2001). Initially created as a paper/pencil, 75-item test on a 6-point scale administered to approximately 350 students in a teacher education program, this instrument underwent numerous refinement stages before being made accessible online as a 45-item, 5-point scale with automatic scoring and profiling.

The results of the paper/pencil version, administered to more than 1200 respondents, were not significantly different from online version, administered to approximately 3000 respondents (Pratt, et al., 2001). The instrument includes basic demographic questions to examine correlations between TPI scores; e.g., gender bias, work-related experience, work location, and ethnicity.

2.1.2.10 Thermal and Transport Science Inventory (TTSI)

The concepts in the Thermal and Transport Science Inventory (TTSI) were established through a Delphi study conducted with a group of 30 faculty experts in engineering and engineering textbook authors (Olds, et al., 2004). Experts were defined as, “people with at least five years of teaching experience and involvement in the engineering education community” (Olds, et al., 2004, p. 2). The experts helped the researchers to identify major concepts. After the researchers selected 28 concepts, the experts ranked each concept on the degree of student understanding expectation and on the level of importance placed on knowing the concept. Based on this feedback, ten key items were selected to inform the construction of multiple-choice questions.

In addition to the Delphi study, think aloud interview sessions were held with six student volunteers to test the TTSI questions for readability and to collect data that could be used to inform the construction of distracters for the multiple-choice response options.
True to many of the concept inventories, the questions were adjusted prior to moving to the instrument testing phase.

### 2.2 Reliability and Validity

Reliability and validity are two primary considerations during instrument development. Concept inventories use various types of reliability measures. Basically, to be considered reliable, repeated instrument administrations should yield consistent outcomes. For concept inventories using classical test theory, reliability is primarily established during the data analysis phase using the Cronbach’s alpha reliability coefficient. Developing an instrument with a small measurement error is foundational for establishing validity.

Describing reliability considerations for the construction of the Statistics Concept Inventory, Allen notes, “There are several types of reliability: test-retest measures answer stability on repeated administrations; alternative forms requires subjects to take two similar tests on the same subject; internal consistency is based on inter-item correlations and describes the extent to which the test measures a single attribute (e.g., statistical knowledge)” (2006, p. 15). Internal consistency is a common measure because it requires only one test administration and it is concerned with the consistency across test items within an instrument. While an earlier study determined that a test is considered reliable if the Cronbach’s alpha estimate is above 0.80 (Nunnally & Bernstein, 1994); other concept inventory sources note that estimates in the 0.50 to 0.60 range are acceptable for group level analysis (Maloney, et al., 2001).

In CTT, increasing the number of items within the instrument is known to increase reliability; however, the subsequent lengthening of the allotted completion times may lead to negative consequences, e.g., incomplete responses, increased weariness and decreased focus. A concept inventory emphasizes a balance between test length and quality of item selection during the instrument design process. Generally speaking, the larger the number of participants in a study the more stable the statistics and the more representative the outcomes are of that population.

Instrument design and test theory models inform the reliability measures decision processes, e.g., classical test theory benefits from a large number of test items. It should
be noted that some control for internal reliability may be met through the instrument
design, e.g., the length of the instrument and the number of participants. However, there
are other factors that may reduce internal reliability that are more difficult, if not
impossible to control, e.g., individual motivation or environmental situations.

Does the instrument measure what it is intended to measure? The instrument
design will need measures in place to insure that the instrument is measuring the intended
concepts. Because there are different types of validity, planning for validity is central to
the instrument construction process. Discussing steps for constructing a concept
inventory, Richardson, Steif, Morgan, and Dantzler support seeking expert opinion to
strengthen the instrument’s validity (Richardson, Steif, Morgan, & Dantzler, 2003). The
Delphi study approach is foundational in several of the concept inventories (Danielson,

Validity may also be improved by the basic format of the instrument. For
example, construct validity in concept inventories may be seen in the inclusion of
subtopics and the assignment of difficulty levels. Planning for a combination of data
collection points and measurement analysis will improve the degree of validity. To
provide evidence of validity, Flowers and Algozzine conducted a factor analysis for
construct validity, a BTCEI correlation to a task performance-based activity was used for
concurrent validity study, and a pretest/instruction/posttest technique was used to
measure internal consistency (2000).

Concept inventories designed to capture differences in exposure to content
domain information by default need to have content validity. For example, it would be
expected for someone with emergent literacy domain training to score higher on the
instrument than a person new to the content area. Testing the CI with a wide range of
expected core knowledge levels would be supportive of establishing content (or face)
validity. By organizing common misconceptions related to statics concepts by categories
associated with prior experience or exposure to statics, Stief and Dantzler, for example,
tried to use the distracters in the instrument’s multiple-choice options to strengthen
content validity (2005). Other studies emphasize the importance of consulting experts to
help establish that the instrument is measuring the intended concepts (Danielson, 2004;
Olds, et al., 2004).
Incorporating an interview phase when developing the concept inventory instrument is another common method used to check for construct validity. Again, the idea is to reduce the chances that the wording of the instrument items may be misunderstood and, as such, threaten validity. Attending to the sensitivity of group differences will strengthen content validity by reducing the influences of potential bias; for example, cultural or language differences.

Concurrent (or criterion-related) validity may be established through the comparisons of CI outcomes to other pre-established forms of assessment. For example, Steif and Dantzler compared the participant’s Statics Concept Inventory outcome to the overall course final grade (2005). The ability to use the participant’s performance on a concept inventory to accurately predict future success may indicate that a level of predictive validity has been established. Predictive validity is a type of criterion-related validity. The ability to predict the likelihood of a student's success rate in an introductory course could expand the use of a concept inventory as an advisory tool. This section has presented the topics of reliability and validity from a classical test theory liner model approach. Under item response theory models, validity and reliability assume different interpretations as each item in the instrument has a mean. The question of validity is tied to Theta or ability so it becomes an issue of reliable or valid for whom!

A review of the earlier concept instrument development literature illustrates that validity and reliability are measures of degree. The development of EL-Capstone involved designing procedures that helped to strengthen these constructs; however, future research will continue to address issues related to validity and reliability.

2.3 Theoretical Considerations

The theoretical framework of two test theories informed the design of EL-Capstone: classical test theory (CTT) and item response theory (IRT). These two theories guided decisions related to format and application of statistical measures. The actual format of EL-Capstone was designed during a pre-participant stage and the instrument had an iterative component to the test-item process. Each of these test theories has data criterion and assumptions; therefore, both test theories were included in the literature review. The goal, when applying CTT or IRT, was to identify test items that reliably
discern the ability levels of emergent literacy concepts that support the psychometric soundness of EL-Capstone. Both CTT and IRT attempt to explain measurement error. It is the underlying assumptions associated with the statistical analysis that highlights the differences in the two theories.

### 2.3.1 Classical Test Theory

Classical test theory (CTT) has rules or assumptions that had the potential to be problematic for the EL-Capstone initial design phase. The model of measurement error is based on the correlation coefficient and explains the errors in terms of an observed correlation and a true correlation (e.g., the Spearman-Brown formula or Cronbach's alpha). For example, the CTT requires that the test standard error of measurement is consistent across the entire population. Stability of test reliability increases with the greater number of test items included in the instrument and the greater the total number of people in the test population sample. Concept inventories by design are not lengthy. CTT is stringent with format or item alterations from the original instrument properties. The EL-Capstone length and format was subject to change based on the participant design phase of this study. Tracey acknowledges strengths in CTT applications but emphasizes "all standard errors are assumed to be unconditional in that every score has the same standard error" (2010, p. 3).

Conducting an analysis of an instrument is an attempt to collect information related to the usefulness of a test item and the relationship to how a test item relates or performs in conjunction with the other test items. Basic descriptive statistics offers some information related to usefulness. For example, one would want to look at the means and standard deviations at the individual test-item level and at the overall test level.

The usefulness of a low variance item is extremely limited. An example of low variance would be a test item with all of the participants scoring high. There may be many reasons why the item scored high, but there is just not much information available for further analysis. What is useful is an item with high variance. In this situation, the information gathered indicates something is making a difference in this question. Additional analysis may parse out more information. Perhaps the test item is poorly constructed and a high amount of guessing occurred. High variance may also result from
the influence of a latent variable related to levels of expertise in the subject matter. In EL-Capstone, there was an expectation for the individual test items to have high variability (scores spread across the range of possible results) and to have a mean clustered near the center point of distribution.

CTT has a descriptive statistical term called an item’s difficulty level. The difficulty level, commonly referred to as the pass (p) value, indicates how many of the participants passed an individual test question. If all of the participants answered a test item correctly, that item would have a p value of 1.00. The inverse situation, with a test item incorrectly answered by all participants, would have a p value of 0.00. Summing an individual’s total p values provides a composite score. This may be very useful depending on what the item is designed to measure.

In future applications of EL-Capstone, all of the practitioners in a particular testing site may be highly trained in the emergent literacy concepts being measured and a high p value for all test items would be expected. Using the p value in the initial design phase was thought to assist in the flagging of test items considered too easy or too difficult. The p value may quickly differentiate a person's level of expertise, albeit with somewhat limited information.

CTT uses the p values to determine discrimination indices (D). Like the earlier descriptive measures, test items with high levels of variance and with p values on the mid-range will rank the ideal D values. The D value is calculated through some basic grouping techniques that separate the highest and lowest overall test score performers. The p values are then calculated at the newly created group levels and the difference for each test item. The difference between the groups for each item is the test-item D value.

Calculating the D value for dichotomous items (test items with a right/wrong choice) discriminates the likelihood that a person will pass a particular test-item based on that person’s overall score. In the EL-Capstone, a higher D value on a test item should help discriminate a person’s likelihood of being able to answer correctly based on a greater level of emergent literacy concept awareness. Item analysis assesses the ability of the individual test items to discriminate the various levels of conceptual understanding the person has of emergent literacy concepts. Using CTT plot items curves provides a
greater resolution of discrimination by visually mapping the p values of an individual’s
test items to the total score percentiles.

A few caveats to using CTT. It is sensitive to instrument design modifications
(e.g., changing the number of response options). Kline warns “it is unwise to change the
scales from their original format because the properties of the new instrument are not
known” (2005, p. 94). If using a dichotomous format (e.g., correct or incorrect), CTT
restricts factor analysis and weakens the ability to establish validity (Kline, 2005).
Perhaps the most important limitation of CTT to highlight is the model’s inability to
allow the researcher to separate the sample population characteristics from the test-item
characteristics (Henard, 2000). CTT is considered to be sample dependent and to be
overall test centric rather than being test-item oriented. Hambleton and Slater highlight
that “a second well-known shortcoming of classical test theory is that comparisons of
examinees on the test score scale are limited to situations where examinees are
administered the same (or parallel) tests” (1997, p. 22). Under the CTT model, the
difficulty level of a test instrument may result in the same examinee having varying test
scores (Yu, 2007).

Classical test theory is known for simplicity and flexibility to be applied across a
wide range of test environments (Fan, 1998; Hambleton & Rogers, 1991). Flores
condenses the definition of CTT as a “test theory that decomposes observed score (X)
into the sum of true score (T) and random error score (E)” (2007, p. 6). In CTT, Courville
posits that “item discrimination statistics focus not on how many people correctly answer
an item, but on whether the correct people get the item right or wrong” (2004, p. 4). The
basic CTT tenet is to determine the ratio (p-value) of an individual’s success rate on
passing a test-item to the total number of test takers. So ideally, if the true score equaled
the observed score (an individual’s score) then the test would be considered to be
perfectly reliable.

2.3.2 Item Response Theory

Item response theory (IRT) reveals greater depths of information related to the
response tendencies. In essence, using the IRT approach allows an examination of the
relationship of how each item is being responded to differently at each level of the
construct; i.e., in this case the level of emergent literacy expertise. The literature references item response theory with numerous psychometric terms. In addition to being referenced as the modern test theory, Hambleton and Slater note that IRT terms include “the Rasch model, the three-parameter logistic model, latent trait theory, item response theory, latent ability, item characteristic curves, [and] computer adaptive testing” (1997, p. 21). IRT has been evolving for some time now into accepted models for working with latent trait factors. In presenting a brief history of IRT, Bock points to Thurstone's 1925 paper on scaling methods that has plot features similar to modern IRT (2005).

The IRT models allow for information on both the responder’s expertise level and the test-item’s parameters. Unlike CTT, the IRT model is not dependent on the entire sample population’s expertise level. Thus, IRT allows for the individual level of expertise for a comparison. This would be similar to evaluating the discrimination parameter of a person’s expertise level to other people with similar levels and the likelihood of similar responses on a test-item (Hambleton & Rogers, 1991).

IRT allows for a test administrator to predict an individual respondent's likelihood score based on that individual’s stated expertise level. Because IRT permits a researcher to parse information related to the individual’s performance in relationship to an underlying latent trait (also known as the ability parameter), this model is touted as an alternative test theory that addresses CTT limitations (Cantrell, 1999; Courville, 2004; Henard, 2000).

The ability parameter, theta (θ), is the IRT approach to the same relationship that the CTT establishes through the true score and the observed score. The true score in IRT is based on the latent trait rather than on the test. It is the underlying theoretical assumptions that generate the flexibility in the IRT parameters. IRT posits that the instruments (θ) are not bound to the individual test items and consequently, test items may be added or dropped without generating statistical errors. Barring sampling error, the item difficulty statistics under an IRT model would not change if different participants were used to create the group population (Henard, 2000).

IRT is a mathematical model and generates item characteristic curve plots. The curves represent information related to the probability that a correct response will change as a function of the instrument’s latent trait (e.g., individual differences in emergent
literacy knowledge). Henard contends that the various model distinctions are based on how the “models differ principally in the mathematical form of the ICC and the number of parameters specified in the model” (2000, p. 70).

Under the IRT model, if the EL-Capstone was a dichotomously scored test, there would be three potential models available. The function of the one-parameter (1P) model (the simplest model) holds the item discrimination constant while the item difficulty is permitted to vary. The difficulty parameter is also referenced in the literature as the B parameter and the threshold parameter (Yu, 2007). With the item discrimination held constant, the ICC plot has a unique characteristic of no two ICCs crossing over each other (Yu, 2007). The Rasch model has become a popular one-parameter model (Courville, 2004; Henard, 2000).

If the item discrimination (also known as the A parameter) and item difficulty (B parameter) are permitted to vary across test items, then the two-parameter (2P) model is used. The ICC plot may have some cross over as the 2P shows the discrimination in the relationship of the changes in levels of proficiency performance among the examinees (Yu, 2007).

The three-parameter (3P) model is recommended for multiple-choice and true-false test-item formats since a pseudo-guessing parameter is added (Courville, 2004). Forero and Maydeu-Olivares posit that “in principle, IRT models are better suited than factor analysis for modeling the categorical ordered data arising from the application of rating scales” (2009, p. 275). This added parameter adds the probability that a correct answer will be given for examinees with similar characteristics. Yu contends that, “When the guessing parameter is taken into account, it shows that in many items, even if the examinee does not know anything about the subject matter (θ=-5), he or she can still have some chances (p>0) to get the right answer” (2007, p. 10).

Applying IRT heightens the importance of meeting the assumptions of unidimensionality (only one dominant ability trait is being measured) and local independence. Courville describes local independence as an assumption that “excluding ability, there is no relationship between the test items and the examinee’s responses” (2004, p. 7). Meeting the assumptions of IRT is more stringent than the assumptions of CTT where there is not an expectation of demonstrating the stated assumptions. Yu offers
the following IRT model selection guideline, stating “the more parameters we want to estimate, the more subjects we need in computing. If there are sample size constraints, it is advisable to use a 1P model to conduct test construction and use a 3P as a diagnostic tool only” (2007, p. 10).

The IRT model allows for the test-item attributes (e.g., the difficult parameter) and the individual’s attributes (e.g., theta) to be displayed on the same scale. This scale is known as the logit scale. This information may also be presented in an Item-Person Map (IPM) to provide a quick over-view of the relationship of the individual and the test items. Brannick contends, “IRT really is superior to classical test theory with respect to the idea of local error, local reliability, and local information” (2009, p. 10). Reeve and Mâsse posit that it does not mean abandoning one test theory for another but "rather, IRT methods should complement CTT methods" (2004, p. 271).

2.3.3 Contrasting Classical Test Theory and Item Response Theory

As demonstrated in the preceding sections, these two test theories attempt to explain measurement error and they both have stated data criterions and assumptions. In summarizing CTT and IRT, Reeve and Mâsse note that these two theories differ in the measurement error and precision, the impact of the scale length and instrument format, and the invariance of the item parameters and scale scores (Reeve & Mâsse, 2004).

The psychometric soundness of EL-Capstone was thought to potentially require a combination of these theories to identify the test items that reliably discriminate ability levels of emergent literacy concepts. For example, using the CTT approach, the reliability coefficient and the standard error measurement value would be the same. All persons correctly answering 80 percent of the EL-Capstone would all have been seen as equal score outcomes. This information was seen as being helpful in providing an initial picture of the inventory total scores. If there is little variance among scores, CTT will reveal that information simplistically. For information related to the test-item level, under the IRT model, the information curve would be able to discern differences since the measures of precision are estimated individually and they may vary across scores. In other words, the test-item characteristics of EL-Capstone for participants with identical scores under IRT would be controlled.
The length of the inventory as it relates to reliability has been highlighted in previous sections of this review. It is important to note that CTT and IRT had the potential to impact the construction of the test items at a deeper theoretical level. For example, under the CTT model, adding in numerous similar questions that address an identical concept may have been encouraged for validity; however, addressing redundancy under the IRT model, Reeve and Mâsse state "redundant items are discouraged and may violate the assumptions of local independence" (2004, p. 257).

Another example of CTT's and IRT's theoretical implications on EL-Capstone may be seen in the format design decisions. While many of the inventories have used a pure multiple-choice format, other multiple-choice inventories at some stage of their development have included some open-ended questions (Maloney, et al., 2001) or demographic items (Hott, 2006). A 6-point scale, used in the Teaching Perspectives Inventory, is an example of a successful alternative to a multiple-choice inventory format (Pratt, et al., 2001). IRT easily accommodates mixed item formats. In some test instrument situations, CTT has been extended over the years to find workarounds that will handle unequal test-item categories, e.g., computing z scores. Embretson and Reise note "this [workaround] approach has the disadvantage, however, of depending on the item means and standard deviations in a particular sample" (2000, p. 32). In summarizing CTT and IRT measurement differences, Embretson and Reise state that all of the differences are not equally important and they note "the impact of test length on measurement error... applies to all tests. ... On the other hand, the rule about mixed item formats applies to many, but not all, tests" (2000, p. 39). DeVellis acknowledges reliability as a fundamental issue and highlights that scale reliability is "the proportion of variance attributable to the true score of the latent variable" (2003, p. 27).

2.4 Summary

Concept inventories are reliable instruments that help to establish baselines within a narrow domain. EL-Capstone will provide an instrument for practitioners, educators, policy makers, and other parties interested in gathering emergent literacy core-knowledge baselines of adults working with young children and emergent literacy development.
A multiple-choice item format, within a usually narrow content area that includes embedded subset topics, is a common concept inventory attribute. The instrument development process does have a wide range of developmental variations that have been used successfully. Some instruments, like the Statistics Concept Inventory, start with the identification of the central concepts. Other successful inventories identify pervasive misconceptions as an earlier phase; e.g., Streveler, Olds, Miller, and Nelson’s inventory in thermal and transport science (2003).

There are usually several questions ranging in difficulty levels that help measure each concept. While concept inventories do vary in length and in the expected completion times, a general trend is to design an instrument that may be administered in approximately 30 minutes.

If an instrument’s multiple-choice answers contain embedded common misconceptions or detractors, the methods used to construct these options often include conducting focus groups, personal interviews, and/or administering written open-ended essay questions with the intended sample populations. Some instruments have experts write the answers based on previous teaching or research experience related to the concept.

Common to all concept inventories included in this literature review is an iterative process. The instrument is pre-tested and adjustments are made as needed to improve the reliability (repeatability of outcomes) and validity of the instrument. As previously mentioned, the validity of an instrument may be established in several ways, e.g., domain experts involved in the item development process, and interviews.

A concept inventory has inherent conflicts. Encouraging the dissemination of an instrument across a specific domain contributes to the validity of an instrument. The flip side of encouraging access and use of an instrument may also jeopardize both the validity and the reliability of the instrument. Garvin-Doxas and Klymkowsky illustrate this conflict with an example of possible instructor misuse of the instrument by linking course grades and inventory items as part of the student evaluation process (2008). This, they continue to argue, may lead an instructor to use the questions for the basis of content teaching (even subconsciously) and thus invalidating the results of the instrument.
The El-Capstone initial design goal was to provide a valid and reliable instrument to capture existing capstone levels of persons that are currently working (or may be working in the near future) in environments with birth to pre-k children. Consequently, any supplemental learning of any emergent literacy concept that may occur during the competition of the instrument would be seen as a benefit to the community at large.

Why a concept inventory for emergent literacy? Why not just spend an hour or two and create a quick survey? Would it not have similar outcomes? First, there are the obvious bias factors. One person’s perspective and interpretation of the emergent literacy research would be a limitation. The language or concept terminology in a single author design may, in itself, restrict responses and create misleading distracters in the test item multiple choices. Acceptance of the instrument for future research objectives also places a high demand on the need for the instrument to be based on emergent literacy concepts that are backed by valid and reliable research studies.

Many of the science instruments have a dedicated content domain with a population usually consisting of undergraduates students. The test item questions are agreed upon knowledge concepts and the multiple choice response options may be informed by student misconceptions.

Concept inventories are used to support additional research projects. For example, a national study was conducted to assess the effects of interactive instruction teaching and the learning of introductory astronomy by undergraduates (Prather, Rudolph, Brissenden, & Schlingman, 2009). The study administered the Light and Spectroscopy Concept Inventory (LSCI) in conjunction with an instrument designed to measure the amount of classroom time spent on learner centered interactive instruction. In this example, the concept inventory was administered at 31 institutions as and a posttest instruction instrument to nearly 4000 students.

Like any research study, it is important not to over generalize the applications of an instrument. For example, some concept inventories are used as a subject matter placement test. While the outcome of a concept inventory is able to generate data that may be used to create cut scores that may define participant baseline scores, it is important for that instrument to have previously demonstrated concurrent validity. Henderson (2002) cautions against reliance on the Force Concept Inventory pre-test as a

51
placement test. Henderson’s findings indicate “there is little difference between FCI scores with a test that is given graded versus ungraded” and “giving the FCI as a pre-test does not affect the post-test results” (2002, p. 546).

All of the concept inventories covered by this literature review were developed under the theoretical influences of classical test theory and/or item response theory during various phases of the instrument design or expanded instrument applications. The work of earlier concept inventory development processes informed the research design of the EL-Capstone concept inventory.
CHAPTER 3

RESEARCH DESIGN

The primary goal of this research was to design a valid and reliable community-sensitive tool to support the need of agencies to assess emergent literacy understanding in adults who may be called upon as leaders to promote emergent literacy development of young children. Earlier concept inventory instruments from other domains and test theory informed the development process of this emergent literacy assessment instrument (EL-Capstone).

This research design had three distinct stages: Pre-participant, Participant, and Post-participant. Table 1 outlines the relationship of the research stages to the general task activities.

Table 1

<table>
<thead>
<tr>
<th>Relationship of Research Stages to Task Activity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stage</td>
</tr>
<tr>
<td>Pre-participant</td>
</tr>
<tr>
<td>Participant</td>
</tr>
<tr>
<td>Post-participant</td>
</tr>
</tbody>
</table>

Note. Task key (1) define the content, (2) develop and select the instrument items, (3) informal informant sessions, (4) interview sessions with librarians, (5) instrument administration, and (6) evaluation and final refinement.

The processes used in the development of the EL-Capstone was summarized by six general tasks: (1) define the content, (2) develop and select the instrument items, (3) informal informant sessions, (4) interview sessions with librarians, (5) instrument administration, and (6) evaluation and final refinement. These tasks were not exclusively sequential in nature. For example, an iterative component was anticipated for task 2 since item refinement is central to the design of this instrument. The consulting of informal informants spanned all two stages of the instrument development. All of the consent-
form participant activities were designed with consideration of improving the content validity, internal reliability, and stability of the inventory.

3.1 Stage One: Pre-participant

During the first stage of instrument development, the research was conducted without any recruited consent-form participant involvement. The primary goal of Stage One was to define the scope of the emergent literacy concepts and to generate a large pool of potential items for the first version of the EL-Capstone instrument.

3.1.1 Content Definition

A number of sources were used to inform the scope of emergent literacy concepts; however, the National Early Literacy Panel (NELP) Report was foundational during the initial phase of the EL-Capstone instrument development. NELP (2008a) conducted an intensive literature search of the early literacy publications and evaluated approximately five-hundred research-based articles to inform the basis of their 2008 meta-analysis report. While the definition of the term concept needed to be negotiated during this early phase, the major platform stemmed from the NELP (2008a) report since EL-Capstone is grounded in research-based concepts.

The National Early Literacy Panel was comprised of a wide range of early learning experts (e.g., professors, physicians, and professional development experts). The NELP members' internationally acknowledged expertise within the domain of emergent literacy allowed for the assumption that their 2008 report represents the generally accepted research-based emergent literacy concepts. Even working within the NELP framework, decisions during this content definition phase had to be made concerning format and concepts that were to be included in the EL-Capstone instrument.

The generation of the test pool was informed by the superordinate (sibling) and subordinate (child) concepts identified through a mapping of the NELP (2008a) report. It was presumed that some of the concepts would be overlapping in the literature; therefore, mapping the emerging superordinate (sibling) and subordinate (child) concepts assisted with information management. The emerging structure contained information for the conceptual framework that supported the development of the initial test pool.
Software products were available that met the needs of the concept mapping activity; e.g., The Brain: Visual Information Management (Brain Technologies, 2009). As shown in Figure 2, the Brain5 interface allows for the parent/sibling/child connections to be organized visually.

![Figure 2](image)

*Figure 2.* A basic schematic of the parent/sibling/child relationship connections that may be created using The Brain5 software.

Beyond the Table of Contents of the NELP Report, the actual structure of the document was not known prior to the mapping. The mapping of reading readiness concepts, as defined by the NELP 2008 report, provides an example of a potential parent/sibling/child relationship that may have emerged during the development of EL-Capstone. As shown in Figure 3, the overarching *parent* term in this research was emergent literacy. If an *active thought* (also known as a superordinate/sibling concept) was reading readiness, the *child* connections might be alphabet knowledge, print concepts, vocabulary, memory, and phonological awareness. A *jump* connection allows for the capturing of information that is still closely related to the *active thought* but, none-the-less, it is information that may still be considered to be too distant from the *child*.

It should be noted that since the structure of the NELP 2008 report is not known, flexibility of the concept map will allow for medications of connections. The schematic examples provided in this section still represent the mapping structure capability of the existing software programs.
3.1.2 Draft Instrument Items

The items that collectively constitute the central concepts (both superordinate and subordinate) were considered for inclusion as instrument items. A target of producing approximately two times more than the anticipated number of items required for the instrument was associated with this task. Through the use of test theory in the post-participant stage, the test pool was reduced to the targeted number of instrument test items by identifying the test items with the greatest discrimination ability. Items were also evaluated based on the difficulty parameter.

A common format of a concept inventory is a multiple-choice design with five response options. One correct response is typically accompanied by four incorrect distracter response options; however, concept inventories have used alternative formats (e.g., a sliding scale agreement format). While future studies may focus on the identification of misconceptions, the primary objective of the EL-Capstone instrument was the development of an instrument that measures research-based concept knowledge levels; therefore, a dichotomous format of true/false was considered. Both the common and the alternative concept inventory formats were considered during this Stage One activity.

3.2 Stage Two: Participant

The second stage of instrument development involved participant/s selected as an informant for informal feedback sessions. The informants were selected for either
subject-matter expertise or for their understanding of the stake-holder environment. The item pool for the EL-Capstone was reviewed by an informant in order to generate general instrument feedback related to missing concepts, wording suggestions, or other test-item recommendations.

Two consent form participant groups were recruited to perform distinct activities for the data collection and analysis portion of this study. The first group of consent form participants, librarians and/or early literacy professionals (ELPs) with connections to public libraries, completed a beta version of the EL-Capstone. These participants were asked to provide feedback related to the readability or the clarity of the test pool items. The librarians and ELPs were also asked for test item inclusion recommendations and for test item modification suggestions. One-on-one interviews with the librarian and ELP participants supported this activity.

Anonymous, heterogeneous adults, who participated in an online administration of the EL-Capstone, were the second group of consent form participants. The data collected through this online administration was not used to draw conclusions about the emergent literacy knowledge levels related to any population group or sub-group. These anonymous participants generated test-item responses that allowed the use of test theory analysis. This analysis was used to inform the final test item selection out of the larger test pool.

During all stages of the EL-Capstone instrument development, the selection of participants was equitable and did not knowingly include vulnerable populations. Since the research tasks and objectives across the Stage Two were distinct, a unique participant consent form, approved through the Florida State University Internal Review Board, was administered for the interview sessions and for the online administration of EL-Capstone.

3.2.1 Cognitive Interviews

Paradigms of cognitive interviewing practices vary. For the purpose of this study, a cognitive interview session held with librarians and ELPs was a process in which additional information about the beta version of EL-Capstone item responses was collected for use in helping to evaluate the instrument and to select the items.
Elaborations of a participant’s interpretation of the instrument items, readability, and suggested alternative wording were the primary goals.

While the EL-Capstone’s purposive outcome was to measure core emergent literacy knowledge-levels of practitioners, such measurement was not an objective during this stage of instrument development. The data collected informed the readability or clarity of the test-item wording and the validity and reliability factors. The information obtained through the interviews was reviewed for possible EL-Capstone item revisions. For reiteration, the primary goal of this activity was to ensure the instrument test items were measuring the intended goals and to check for readability.

3.2.1.1 Practitioner Participants.

Ten librarian practitioners or ELPs were recruited to participate in one-on-one interviews. Participant frequency was one and the duration was estimated to be approximately 45 minutes. Based on the typical number of participants used in earlier concept inventory studies, the maximum number of librarian and/or ELP participants was ten. The Librarian consent form was used to obtain consent to participate.

3.2.1.2 Procedure.

Each of the librarian/ELP volunteers participated in a one-on-one interview with the researcher. At this stage of development, the beta version of EL-Capstone reflected a larger number of test pool items. Feedback interview questions interspersed with the EL-Capstone test items assisted in measuring the instrument’s intended outcomes. For example, the participants were asked, What do you think this question is asking?, What suggestions do you have to improve the readability of the question?, and other similar questions. Additional comments and overall suggestions to improve the test items were solicited at the end of each interview session. The collected librarian/ELP interview data informed possible instrument revisions.

3.2.1.3 Data Analysis.

The major analysis goal was to combine the qualitative data and to look for themes that may have indicated a need for possible instrument revisions. The goal was to weigh each test item on the ability to measure a stated concept, the quality of the question, difficulty level, and to solicit alternative wording or concept suggestions from
emergent literacy experts. Analysis on the embedded questionnaire was used to inform revisions and test item inclusions for the EL-Capstone.

3.2.1.4 Instrument Refinement.

The instrument was refined as needed according to the librarian/ELP feedback collected through the cognitive interview phase. A revised version of the instrument was developed for the online administration of EL-Capstone task. The instrument item pool at this stage of development still contained more items than the targeted end product of less than 30 items. The targeted EL-Capstone length is influenced by previous concept inventories.

3.2.2 Instrument Administration

The post-librarian/ELP interview version of the EL-Capstone was administered online to anonymous adult participants. It was anticipated that this version of the instrument would have one and a half to two times the number of targeted instrument items. The instrument administration task was used to collect data for additional test theory analysis and subsequent instrument refinement as needed. This information informed the selection of the highest discrimination items that measured the intended goals and it assisted in reducing the test-item length of the instrument.

3.2.2.1 Adult Participants.

Adult participants, over the age of 18, were recruited through various fielded-related resources (e.g., list-serves, professional organizations). A recruited participant goal of 100-300 respondents was large enough to represent adults with different emergent literacy experience levels. The wide range goal of possible respondents reflects the differences of CTT and IRT applicable test theory model assumptions related to the size of n.

3.2.2.2 Procedure.

The EL-Capstone was released online to collect test item responses from adult respondents. Given the potential of a large number of items in the test pool, there was a potential, at this stage of development for the instrument to be released in subsets or in multiple forms in order to minimize participation attrition. The University Internal
Review Board approval of a maximum of 1000 adult participants, for this distinct task, accommodated for the flexibility in design progression decisions.

The continuing goal of this procedure was to eliminate poor performing items while retaining sufficient items that would discriminate the ability trait (i.e., emergent literacy knowledge). The analysis of this data was conducted in the post-participant stage. The data was collected to identify the performance of the items. Items that all respondents got correct or incorrect deliver low levels of discrimination and as such were flagged as potentially low performing items.

3.3 Stage Three: Post-participant

No consent-form participants were involved in Stage Three tasks and no prior participant identifying information was included in the dissemination of this research. The statistical analysis was determined by the actual design of the instrument (e.g., dichotomous test items) and by the best-fit model associated with classical test theory or item response theory. The appropriate test theory data analysis methods were affected by the number of respondents of the online administration of EL-Capstone and the data fit to the model assumptions. For example, classical test theory (CTT) is appropriate for approximately 100-150 respondents while Item Response Theory (IRT) needs approximately 250-300 respondents.

The instrument was refined as needed according to the appropriate test theory analysis. The primary goal of using either CTT or IRT was to retain the test pool items that most strongly measured the intended goals of this instrument (i.e., emergent literacy core knowledge concepts).

After a final evaluation of all of the participant feedback and appropriate test analysis, the EL-Capstone will require additional inventory administrations in future studies to refine the instrument. Future projects may expand the initial application to include the assessment of learning outcomes, the evaluation of teaching methods, and the informing of curriculum design.
CHAPTER 4
PROCESS & RESULTS

This chapter reports on the decisions made during the construction of the instrument and on the Stage 2: Participant data results. The information has been reported in three sections to mirror the three distinct stages; i.e., Pre-participant, Participant, and Post-participant. The analysis was on the construction of the EL-Capstone instrument and does not include any inferential statistics regarding the sample population.

4.1 Stage One: Pre-participant

The primary goal of Stage One was to define the scope of the emergent literacy concepts and to generate a large pool of potential items for the first version of the EL-Capstone instrument. These activities did not have recruited consent-form participant involvement.

4.1.1 Concept Definition

The initial activity was to define the boundaries of what is considered to be an early literacy concept. The early literacy literature targets the importance of a clear definition of early literacy; however, this discussion tends to focus on a broad verses narrow definition and on associated behaviors that may be assessed in young children to identify levels of early literacy knowledge (Mason & Stewart, 1990; Sénéchal, LeFevre, Smith-Chant, & Colton, 2001; Whitehurst & Lonigan, 1998).

A broad view of early literacy includes both conceptual and procedural knowledge. Knowing properties of a book, e.g., right side up, text flow direction, is an example of a conceptual knowledge concept. An example of a procedural knowledge concept is a child singing the letters of the alphabet does not necessarily have ownership of the concept of letters as symbols. A narrow view tends to focus on the more observable procedural knowledge. Translating these definitional approaches to adult early literacy concepts, EL-Capstone incorporates the broad view and includes both conceptual and procedural knowledge statements. In creating an instrument to measure early literacy core knowledge, the term knowledge was used broadly to measure what a person believes
to be true about basic early literacy concepts. This definition is broader than the term *knowledge* that refers to justified, universally accepted, absolute truths.

Professional development literature for early childhood and preschool teachers informed the decisions related to defining the scope of the definition of early literacy. Professional development often adopts a broad view and prescribes "accumulating a desirable set if knowledge and skills" (Dall'Alba & Sandberg, 2006, p. 386). The literature from teacher professional development guides contains connections of teacher knowledge to teacher skills; however, these guides often move outside the scope of early literacy skills to include later conventional literacy development classroom activities and screening guidelines (Alliance, 2000).

Self-reporting surveys within the teaching domain tend to be too narrow in the early literacy practices with an emphasis on phonological awareness, phonics, and syllabication knowledge (Craft Al-Hazza, Fleener, & Hager, 2008; Cunningham, Perry, Stanovich, & Stanovich, 2004; Spear-Swerling, 2007; Spear-Swerling, Brucker, & Alfano, 2005). There are additional teacher scales, e.g., *Literacy Instruction Knowledge Scales*, that include classroom observations and environment measures (Reutzel et al., 2007; Wolfersberger, Reutzel, Sudweeks, & Fawson, 2004). Surveys that attempt to capture the home/family environment EL influences but are not actually addressing the ownership of EL concepts by the care-givers (Constantine, 2004). Therefore, while the classroom teacher literature is informative, this expanded perspective was not a good fit for defining the scope of a concept for the EL-Capstone tool. Discussing scale development, DeVellis (1991) writes:

> Even if there is no available theory to guide the investigations they must lay out their own conceptual formulations prior to trying to operationalize them. In essence, they must specify at least a tentative theoretical model that will serve as a guide to scale development. (p52)

The boundaries of the terms emergent literacy, early literacy, and early learning are not well-defined. Broader definitions integrate early childhood brain development research, family and social influences, and the role of play in the early learning domain. As a concept inventory, EL-Capstone assesses key concepts in the domain. While the decision to restrict the scope of the instrument to the variables addressed in the *Developing Early*
Literacy, National Early Literacy Panel 2008 Report might be considered by some to be a conservative approach, this meta-analysis document provided a concrete reference point for bounding the scope of the early literacy evidence based concepts and to define the terms. Downing and Haladyna posit, "No other issue is as critical, in the earliest stages of developing effective tests, as in delineating the content domain to be sampled..." (2006, p.7). Subsequently, statements directly related to brain research and play were not included in the scope of this instrument.

With a decision made to restrict the concepts to those contained within the Developing Early Literacy document, the next step was to map out the NELP Report to visually connect the structure of the report to potential concept inventory items. The SciPlore mapping software was chosen for the task of mapping out the content of the Developing Early Literacy, NELP 2008 report. Developed by researchers at the Otte Von Guericke Universitat Magdeburg (Germany), SciPlore is a free mapping tool that includes bibliographic support (Beel, Gipp, & Müller, 2009). SciPlore was the ideal mapping software for this task. The tools are powerful and the program is aligned with the developers' stated criteria to provide tools that allow researchers to focus on the task at hand; i.e., evaluating scientific publications, without being distracted by the technology tools needed to complete the work (Beel, Gipp, & Müller, 2009).

A portable document format (pdf) of the Developing Early Literacy, NELP Report was attached to the central NELP Report node in SciPlore. Using highlighting tools and bookmarking tools available in Adobe Acrobat, the report text was marked for important findings and for information that may support the development of potential test items. Bookmarking in a pdf file creates an organized hypertext markup of the information that permits hyper linking to existing document text or to thoughts connected to the information. The latter is done through the creation of personal note bookmarks.

The structure of the report emerged through the new child or sibling nodes. Figure 4 provides an example of the chapter mapping process of the Developing Early Literacy, NELP 2008 report.
Figure 4. An example of the mapping process of the *Developing Early Literacy*, NELP 2008 report.

The mapping process drilled deeper and deeper into the report structure. Figure 5 represents how chapter 2 expanded into the categories of analysis and visually supports the systematic processes employed by the NELP report researchers.

Figure 5. An example of the chapter 2 mapping process of *Developing Early Literacy*, NELP 2008 report.

It should be noted that the *Developing Early Literacy: Report of the National Early Literacy Panel* is an extremely dense synthesis of the early literacy experimental and quasi-experimental research that met the inclusion methodology laid out by the NELP panel. The information extracted from the mapping of the document's structure incorporates theoretical underpinnings associated with the NELP Panel's stated assumptions. For example, the visual representation of the predictive relationship between variables and decoding skill shown in Figure 6 are bounded by the variables included in the studies that met the NELP inclusion criteria.
Figure 6. An example of the Developing Early Literacy, NELP 2008 report's chapter 2: Meta Analyses mapping of the predictive relationship between variables and decoding dependent on the age assessment.

The expansion of the meta analyses conducted under the primary analyses reported in Chapter 2: Identification of Children's Skills and Abilities Linked to Later Outcomes in Reading, Writing, and Spelling was subject to the overall NELP assumptions and inclusion decisions. Familiarity with the methods related to the systematic search, as detailed in chapter 1 of the NELP Report, is an important factor. For example, the statistical weighting of the effect size was influenced by the panel's criteria decisions to confine possible predictive variables to those that were represented in at least three studies for inclusion in the chapter 2 primary analyses portion of the report. One might also need to understand term definitions; e.g., the term children refers to young children age birth through kindergarten or five years or that the term decoding, if not otherwise specified (NOS), refers to the "use of symbol-sound relations to verbalize real words, pronounceable nonwords, or both" (NELP, 2009. p. 42).
In addition to nodes, the SCiPlore note tools (designated by the symbol 📝) served as an additional method to reference key findings or specific information that might be needed to construct EL-Capstone test items. For example, a note associated with the strong relationship child node was a quote from page 57 of the NELP report; "For these analyses, average correlations that were 0.50 or larger (i.e., the predictor variable explains at least 25 percent of the variance in the outcome variable) were designated as strong relationships" (NELP, 2009, p. 57). An indication that the predictor variables are listed in order of "the strength of correlation with later decoding measures" is another example of a note associated with the strong relationship child node as shown in Figure 6 (NELP, 2009, p. 59).

Unlike a person's weight, for example, that may be measured with standard scales, the concept of early literacy knowledge is a latent variable and it cannot be measured directly. The unidimensionality of the EL-Capstone's early literacy construct was achieved through the construction of indicator variables. The development, administration, and analysis of a large pool of potential instrument items assisted in the identification of the indicator variables that may serve to measure the latent variable, early literacy knowledge.

### 4.1.2 Draft Instrument Items

The construction of a large pool of potential instrument items began with the creation of a three-category classification structure for the types of indicator variable items. As indicated in Figure 7, the instrument items fell into three general categories: predictive, definition, and activity-based. These were not mutually exclusive categories. For example, an instrument item may measure both a definition and an activity.

These three categories cross various content levels of understanding. For example, a statement may address a basic knowledge concept, i.e., definition, while other instrument items may ask for application of that basic definition knowledge. For example, the statement, *reading stories or poems with rhymes or different sounds to a child helps develop future decoding skills*, measures an understanding of the definition of phonological awareness and assesses the transfer of this definition to a predictive value of an activity to later conventional literacy outcomes.
Figure 7. Predictive, definition, and activity-based items comprise the indicator variable categories.

The potential instrument items were sorted into two general headings, i.e., NELP-generated (N) and Alternative-input (A). The NELP-generated items were directly tied to the NELP Report mapping. Two potential "N" items, demonstrate the relationship of "N" items to the NELP Report (2008):

- Based on existing research-based studies, there is no evidence on which to conclude that children’s ethnicity moderates the positive effects of code-focused interventions, and
- Based on existing research-based studies, there is no evidence on which to conclude that population density moderates the positive effects of code-focused interventions (p. 117).

The "A" statements were generated from the early literacy concepts addressed in the NELP Report; however, they were translated into more generic laymen terms and represent early literacy concepts as they are typically presented in early literacy training programs. For example, A30 (Storytelling has been shown to be a strong predictor in later reading comprehension) is not directly linked to text quoted from the NELP report.

Research demonstration on the correlation between early literacy knowledge levels of adults working in environments with expectations of supporting early literacy development and maximum child conventional literacy gains have not been established. Without prior early literacy adult concept knowledge baselines, creating an instrument across a continuum of early literacy core knowledge has to account for the variance in the
terminology used in the field when addressing these concepts. Noticeable differences in the vocabulary may be observed between "N" items and "A" items. Planned feedback during the Participant Stage provided an opportunity to receive user input related to the variance in vocabulary factors.

Operationally defined term differences are not unique to early literacy concepts. In the construction of the Signals and Systems Concept Inventory (SSCI), the researchers encountered difficulties with even well-defined concepts, e.g., frequency. They noted that frequency may be referenced in the literature as radian frequency or by the term hertz (Wage et al., 2005). To address vocabulary issues, the researchers simply made a choice. Wage et al. explained that there was "no intrinsic reason for the SSCI to use radian frequency, and future versions could easily be written using different notation" (2005, p.450).

Existing concept inventories range from 20 to 48 items; however, there has been some discussion to reduce the number of items to avoid subject fatigue during the testing process (Libarkin, 2008). The length of these earlier concept inventories informed the number of items used throughout the EL-Capstone development stages. The actual length of the instrument is still to be informed by the data analysis of the anonymous adult participants' responses to the online EL-Capstone administration in Stage Two: Participant; however, the development length was targeted not to exceed 30 items.

Discussing guidelines for test construction, Massad posits "the true-false or alternative response question is most useful where the desired outcomes are determining the truth or falsity of a statement, distinguishing between fact and opinion..." (2003, p.4). The format design was selected to allow for objective true/false scoring. This item format was chosen because a person taking the instrument would either identify a statement as an evidence-based early literacy statement, and therefore true, or as a false statement. TRUE/FALSE dichotomous scales may also result in less variance and covariance than other formats (DeVellis, 1991).

Primarily, since no previous systematic measures of existing early literacy core knowledge concepts of adults had been established, the use of expanded multiple-choice response options based on common misconceptions was determined by the researcher to not be an appropriate format for this early stage of the construction. After baseline
measures have been established, the more traditional concept inventory expanded multiple-choice response format with embedded misconceptions is a feasible format for future EL-Capstone research projects.

The instrument items were developed to parse out differences in early literacy core knowledge abilities. The construction guidelines set for the development of the indicator variables allowed for information to be gathered across a wide spectrum of early literacy concept levels. Each item in the test bank was given equal weight. An item measuring a level of understanding at the application level was not given greater importance, or a greater correctness value assignment, than a definition knowledge item. For example, an item in the item bank that measures an application of phonological awareness concept was not assigned two points while a correct response on correctly identifying the definition of alphabet knowledge was assigned a correctness value of one. Equal weight does not translate into equal item difficulty as identified through test theory analysis.

As in any typical assessment development, writing objective True/False items, having only one correct answer is the optimal outcome. In an emerging discipline such as early literacy research, or within well-established domains, even researchers with similar epistemological perspectives may interpret the research differently. Some concept inventory developers attempted to address these differences through the use of a Delphi study to help construct item banks. While the use of a panel of experts to identify potential differences may broaden the epistemological perspectives, bias or constrained perspectives still exist just through the natural selection processes involved in establishment of a Delphi study. For the construction of the EL-Capstone item bank, the NELP Report established epistemological perspective was automatically adopted.

An additional item construction challenge was the potential for diverse early literacy knowledge backgrounds. For example, N24 (Shared-reading interventions appear to have no impact on young children’s alphabet knowledge; however, there have been too few studies to provide a reliable estimated impact) may be seen as only partially true by an expert who is aware of studies related to shared reading that focus on print that do have this effect. These factors may possibly be perceived as limitations of the instrument; however, they do still carry value in parsing out differences in the ability
levels. Thornton and Sokoloff posit that, "Such results confirm the value of asking a variety of questions of diverse audiences to probe understanding of particular concepts" (1998, p. 340).

Research is an ongoing process. For the first construction stage of EL-Capstone, the NELP Report provided a common ground timeframe for establishing the early literacy concept inventory items for the item bank. The next level of the indicator variable construction involved participant feedback and the testing of the items in the instrument's item bank.

4.2 Stage Two: Participant

Stage Two is a summary of the informal informant session, the outcome of the face-to-face cognitive interviews (n=10), and the data collection analysis of the anonymous volunteers (total n=824). Since the research tasks and objectives across the Stage Two were distinct, a unique participant consent form, approved through the Florida State University Internal Review Board, was administered for the interview sessions and for the online administration of EL-Capstone.

4.2.1 Informant

The Participant stage began with the selection and contact of an informant. A key informant is often a field expert or person in a privileged position and thus able to provide unique information on a topic (Pauwels & Hardyns, 2009). The EL-Capstone informant was a tenured faculty member at an accredited Master’s in Library & Information Science (MLIS) program in the Southwest. As a recent recipient of an IMLS Laura Bush 21st Century Librarians grant that is targeted to increase the number of MLIS-degreed librarians that are prepared to serve the early literacy needs of young children and the community, this informant was able to address the issues related to early literacy, course development, and public librarian preparedness.

A Skype call facilitated the meeting with the informant. Skype is a software application that allows users to make voice calls over the Internet and to share desktops. Because of the innate concept inventory confidentiality factors, this online format allowed the researcher to share an overview of the NELP concept mapping and potential test-items with the informant without the release of the actual documents. Concept
inventories, like any assessment, fall subject to test fraud. Test fraud can result in threats to instrument validity. Impara and Foster note two major fraud classes, "The first class is cheating (obtaining a score through inappropriate means) and the second is piracy (stealing of items, item pools, or tests)" (2006, p. 93). Suggested strategies to minimize test fraud related to cheating and piracy informed the decision throughout the development process of EL-Capstone.

The Skype call lasted approximately one hour. The discussion was guided by the following questions:

- What misconceptions or missing early literacy concept knowledge gaps were identified that precipitated the decision to seek funds to create specialized early literacy curriculum for public librarians?
- How do you think an untrained early literacy librarian would complete the following statement: The single most important parent/child activity to promote early literacy is to ________________?
- How do you think a trained early literacy librarian would complete the following statement: The single most important parent/child activity to promote early literacy is to ________________?
- Where do you think public librarians go for information on early literacy concepts?

The informant expressed a perceived disconnect between the emerging brain development research and early literacy training. In addition, the informant cautioned that the embedded scientific jargon related to literacy acquisitions by early literacy researchers often necessitates interpretation or sensitization to buzz-words. The informant posited that without a sound early literacy foundation, librarians tend to believe what others bring in and these may not be best practice activities.

The informant suggested that an untrained librarian upon being asked to complete the statement, The single most important parent/child activity to promote early literacy is to ________, would most likely respond just read. In contrast, the informant felt that an early literacy trained librarian might respond, The single most important parent/child activity to promote early literacy is to have multi-sensory interactions with books.
During the discussion related to training opportunities and early literacy preparation programs, the informant expressed a belief that librarians seeking early literacy concepts are likely to visit the American Library Association (ALA), Every Child Read to Read (ECRR), and/or possibly State Library websites for information. Evidence-based research dissemination sites were not suggested by the informant, e.g., National Institute for Literacy or the U.S. Department of Education, Institute of Education (IES), What Works Clearinghouse.

In summary, the informant session was a positive information exchange. The informant offered to support recruitment efforts as the instrument moved from the interview task to the online administration. The feedback was confirming in respect to the instrument's progression to the next task, i.e., cognitive interviews.

4.2.2 Cognitive Interviews

Conducting interviews was the process through which additional information about the beta version of EL-Capstone item responses was collected. The elaborations of a participant's interpretations of the instrument items, readability, and suggested alternative wording informed revisions and the item retention for use in the online administration of EL-Capstone. The data collected was not interpreted to reflect any sample population attributes and was used primarily to collect feedback on the structure of the instrument items.

Ten participants that were either librarian practitioners or early literacy professionals (ELP) were recruited to participate in one-on-one interviews. The participants were recommended by the Principal Investigator (PI) of the IMLS National Leadership Planning Grant, LG 5509009909, Project VIEWS: Valuable Initiatives in Early Learning That Work Successfully. The 10 Pacific Northwest participants represented early literacy programs trainers, early literacy directors, library directors, and front-line practitioners.

The one-on-one interviews were conducted in a face-to-face format with procedures consistent with sound research design in place. The interview participation in this study was voluntary and there was no compensation for participation; however, an expanded understanding of emergent literacy concepts may have been a direct benefit
from participation. The participants were free to not answer any question and were free to withdraw at any time.

Voice recording effects during interview sessions have been well established (Belson, 1967). The level of interview participant feedback precision for the interviews was not high relative to the desired outcomes; therefore, the researcher made a decision not to collect voice recordings during the interviews in order to encourage spontaneity in the participant feedback. Interview note-taking was, however, assisted by a graduate student from a leading Northern Pacific University.

The researcher traveled to each location so the participants were in familiar surroundings. Eight of the interviews were conducted within a public library setting. Due to space restrictions, one of the remaining interviews was conducted in a common area of branch public library. The other remaining interview was held in an EL administrative branch office. All locations provided comfortable environments for the face-to-face interviews. Interview participant frequency with each volunteer was one and the duration was approximately 45-60 minutes. The Florida State University Internal Review Board interview task consent form (Appendix B) was used to obtain consent to participate.

The interview iteration of the instrument contained 68 items. To keep the interview sessions constrained within reasonable time parameters, the instrument pool was divided into three sets. The 45 "N" items, directly linked to the vocabulary construction of the NELP Report, were divided into three sets of 15 items each. The 23 "A" items were divided into three sets of seven statements. The remaining two "A" items, A22 and A23, were included in all of the interview session packet sets. Table 2 denotes the breakdown of item pool used during nine of the interview sessions.

Table 2

<table>
<thead>
<tr>
<th>Packet</th>
<th>&quot;N&quot; unique items</th>
<th>&quot;A&quot; unique items</th>
<th>&quot;A&quot; common items</th>
<th>Total items</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blue</td>
<td>15</td>
<td>7</td>
<td>2</td>
<td>24</td>
</tr>
<tr>
<td>Pink</td>
<td>15</td>
<td>7</td>
<td>2</td>
<td>24</td>
</tr>
<tr>
<td>Green</td>
<td>15</td>
<td>7</td>
<td>2</td>
<td>24</td>
</tr>
</tbody>
</table>
The tenth interview set used during the final interview session contained 24 items. Items A22 and A23 were combined with 15 "N" items and 7 "A" items that had drawn comments during the earlier nine interview sessions. This selection process allowed for additional feedback on any flagged items.

Each item statement was combined with supplemental survey items for the interview activity. Nine 24-page interview packets, i.e., Blue, Pink, and Green, were printed and assembled into individual notebooks. One set of all 68 items was printed to support the assembly of the tenth 24-page interview session packet just prior to the final interview.

Figure 8 illustrates the layout of the statements and the three supplemental survey items. These three supplemental items asked the participant to rank (1) the ability of the item statement to address an early literacy concept, (2) the ability of an adult working with young children to answer the statement correctly, and (3) if the statement is an important concept for adults working with young children to understand.

Please indicate if the following statement is "True or False" based on evidence-based early literacy research (i.e., The National Early Literacy Panel Report).

Statement ID #: N1

Statement: Alphabet knowledge (AK) is the knowledge of letter names or letter sounds associated with printed letters.

☐ True
☐ False

Does this item address an EL-literacy concept?

☐ NO ☐ NO ☐ NO ☐ YES ☐ YES ☐ YES

Will adults working with young children be able to answer this statement correctly?

☐ NO ☐ NO ☐ NO ☐ YES ☐ YES ☐ YES

Is this an important concept for adults working with young children to understand?

☐ NO ☐ NO ☐ NO ☐ YES ☐ YES ☐ YES

Please share any helpful comments or suggestions for the improvement of this item:

Figure 8. Interview activity format of the statements and supplemental survey questions.
Feedback interview questions interspersed with the EL-Capstone test items provided additional information related to the instrument’s intended outcomes. For example, during the interview session, the participant was asked to share helpful comments or suggestions to improve the communication and quality of the item statement.

Similar to the True/False format of the item statements, a basic NO/YES format on a 6 strength scale is used for the three supplemental questions. Figure 9 illustrates the numerical coding system used to record the responses.

![Coding of the supplemental survey questions from the interview activity.](image)

After the interview sessions, the "N" and "A" statements and the three supplemental questions were coded in a Microsoft Office Excel spreadsheet. This information was used to support statement item revisions prior to the launch of the online version. The results of the supplemental questions are in Appendix C. It should be restated that the purpose of the interview sessions was to collect feedback on the wording structure of the statement and not on the collection of the participant scores from the True/False responses.

An illustration of the coding process is shown in Figure 10. A NELP-generated statement, N39, written as a false statement was coded with a correctness value of False (0). Two of the three participants evaluating this statement responded correctly and one person was incorrect and responded that N39 was a true statement. The two participants with a correct response rated N39 with a low score for being an early literacy concept and felt that the average adult would have difficulty answering this statement. In contrast, the participant that held the belief that N39 was a true statement also ranked this item as a high early literacy concept. These three participants ranked statement N39 with a Yes response (4, 6, and 6) to indicate a belief that N39 is an important concept for adults working with young children to know.
Part of the interview participant task was to generate alternative wording for concepts that may be understood in alternative vocabularies. Feedback on the clarity of wording was informative but at times quite polar. For example, some participants expressed the importance of retaining the research jargon. This perspective was accompanied by a belief that parents need to be empowered by the research jargon in order to be equal partners with school administrators and staff during discussions related to early literacy topics. Other participants suggested that the research jargon should be eliminated.

The term *interventions* is an example of the polar feedback with comments that ranged from "a negative term" to "jargon can be empowering." The term *shared-reading* is another example that elicited responses, e.g., "one on one," "reading together," and "Does this mean 2 people reading together?" These examples help to demonstrate some of the difficulties involved in constructing evidence-based concept statements for a wide range of early literacy knowledge populations. This complex domain is accompanied by *it depends* responses. For example, using the term *children* may be too broad. An interview participant commenting on an alternative wording statement related to phonological awareness that included the word *children* (A3), illustrated this point stating "not always true...some children have learning disabilities."

Following the interview sessions, the information collected was used to assist in revisions and eliminations. It should be noted that at this stage of instrument development, while informed by the feedback of the face-to-face interviews (n=10), the elimination and revision of items was not based on test theory analysis.

Two "N" statements tied to the NELP Report that were not used during the online administration of EL-Capstone serve as examples of statement items that were eliminated following the cognitive interviews. These two "N" statements, discussed in an earlier section of this document, are early literacy concepts that are directly linked to page 117.
of the NELP Report; however, after consultation with an early literacy expert, they were
determined not to be core early literacy knowledge concepts, as such, they were
eliminated by the researcher from the pool.

Statements of both high and low research jargon structure moved forward into the
online administration of the instrument pool since there was a lack of agreement among
the participants regarding the value associated with retaining or eliminating it. The 68
items prior to the interview sessions were reduced to 54 items for the online instrument
administration of EL-Capstone.

4.2.2 Instrument Administration

At this stage of development, EL-Capstone still had a larger potential item pool
than the targeted goal of the finished instrument. Based on feedback from the cognitive
interviews and previous concept inventory item recommendations (Libarkin, 2008), a
decision to spiral was made. The remaining 54 items were split into two forms for the
online administration of the instrument. Alternating between multiple forms such that a
proportion of the participants are receiving one form while the other participants are
receiving the other form is referred to as spiralling (Kingston & Dorans, 1984). Kingston
and Dorans posited that the use of spiralling in their experience "... results in the
assignment of equivalent groups of examinees to each subform" (1984, p. 148).

Numerous survey creation programs are available for free, e.g., Survey Monkey
and Google Survey. EL-Capstone has characteristics that are more like a test and less like
a survey questionnaire since the items have an associated correctness value. After
evaluating the trade-offs of the various programs, SurveySelect.Asp was selected for the
survey creation and hosting software. This program was selected primarily because it is
an in-house software at the Florida State University College of Communication and
Information located on a Dell PowerEdge 1950 secure server running Windows 2003
Server Enterprise. The server's 30 gigabytes of storage capability with reliable, scheduled
back-ups through Symantec Backup Exec 2010 software provided a secure location for
collecting the data from anonymous volunteers since no identifiable information
collection during the administration of instrument was included.
The various SurveySelect.Asp software output report options were evaluated to ensure that the necessary data for Test Theory analysis were present. The three data format SurveySelect.Asp options were:

1. User Responses - Export each user's response as a row, with one column per question.
2. Individual Responses - Export each answer as a row, with multiple rows for multiselect questions.
3. SPSS Format - Export each response as a row, with one column per possible answer to each question.

The internal code of the SurveySelect.Asp software had a counter script in place to track the total score for each participant; however, the counter script did not allow for the correctness value to be captured at the individual item level. The total score information was only available in the option 1: User Responses data format; therefore, this was determined to be the best format option for this task. Converting the TRUE/FALSE/NO ANSWER responses during the data analysis phase through a Find/Replace All was added to the Post-participant phase planned activities for processing the data.

A decision was made to break the test item pool down into two smaller clusters during the online administration so as not to discourage volunteer participation in the study. A website was needed to redirect the participants to the actual data collection software since SelectSurvey.Asp did not support adaptive routing. The domain name elcapstone.com was selected for the host website. Figure 11 displays the user interface for the website. A general introduction to the volunteer task was provided along with contact information.

Having the two forms created the need to write a load counter script for the How to volunteer link, which alternated between the two forms, i.e., BC and LC. The use of two forms enabled the division of the instrument item pool into two groups. The script was written in ColdFusion 7 using an SQL Server 2005 database backend. Figure 12 presents the basic script that alternated the participants between the two forms.
Figure 11. User interface for http://www.elcapstone.com running the script to counterbalance the two forms.

```cfspecial
<cfparam name="url.l" default="">
<cfif url.l eq "1" and now() ge '08/31/2010'>
  <cfquery datasource="#ds#" name="getInfo">
    SELECT COUNT(urlVar) AS count1
    FROM urlAction
    WHERE (urlVar = '1')
  </cfquery>
  <cfquery datasource="#ds#" name="getInfo2">
    SELECT COUNT(urlVar) AS count2
    FROM urlAction
    WHERE (urlVar = '2')
  </cfquery>
  <cfif getinfo.count1 le getinfo2.count2>
    <cfquery datasource="#ds#">
      INSERT INTO urlAction (urlVar)
      VALUES ('1')
    </cfquery>
    <cflocation url="http://surveyo.oci.fsu.edu/TakeSurvey.amp?SurveyID=9293589B6c3MI"/>
  </cfelse>
  <cfquery datasource="#ds#">
    INSERT INTO urlAction (urlVar)
    VALUES ('2')
  </cfquery>
</cfif>
<a href="#L-1">Simply click here to participate in EL-Capstone research</a>
```

Figure 12. The basic script that alternates the participants between the two forms.
The database backend was running one table with two fields, i.e., pkid and urlVar. The script logic was simple. After the link, *Simply click here to participate in EL-Capstone research*, was visited by a potential participant, the process got underway. The first survey url link was assigned a value of "1" and the database count began collecting the number of "1" hits. The second url got assigned a value of "2" and the database count began collecting the number of "2" hits. Once underway, every time a user clicked on the participate link, the script looked to the database to see how many counts there were for each value, i.e., "1" and "2." The script compared the value counts. If the count for value "1" was less than or equal to the count of value "2," then the url link for "1" was assigned to the user's website click. If the count for value "1" was greater than value "2," then the script redirected the user to the second survey url.

Once the script delivered the volunteer participant to one of the two forms running through the SurveySelect.ASP software, the interface was consistent. Figure 13 illustrates the first page of the instrument. A page progress bar located in the upper right portion of the body text allowed the user to know how much of the survey had been completed. The "back" button option in the survey software was removed to discourage the user from returning to an earlier item statement; however, the SurveySelect.ASP software did not restrict the user from using the back button option of the web browser.

![User interface for the first page of the instrument.](image)

*Figure 13. User interface for the first page of the instrument.*
The Florida State University Internal Review Board adult consent form (Appendix B) was used to obtain consent to participate in the anonymous online administration of the instrument. As shown in Figure 14, the first question asked for informed consent and was flagged as a mandatory response option, restricting the ability to progress further until a response is entered. A Cancel button was available to allow the anonymous volunteer participants to opt out of the research at any time.

1. Please indicate your agreement with the consent form.
   - Yes
   - No

*Figure 14. First question is a mandatory response option to obtain consent.*

Two optional demographic questions in the introduction portion of the survey were intended to support future analysis. These two demographics captured the gender and a self report of early literacy experience level questions. The breakdown of the 54 item pool into the two alternating forms is highlighted in Table 3. Each form began with an "A" item and alternated back and forth with a "N" item until the thirteenth item. At this point, four "N" items and two "A" items were common anchor items that linked the two forms. These six linking items anchored the two forms during the analysis phase and helped approximate the equivalency of the forms. The remainder alternating "N" and "A" items continued until the forms concluded with two "N" items.

Placement of the linking anchor items in the center of the forms was intended to reduce any testing ordering effects that may occur as a participant begins or ends the instrument; e.g., fatigue effects as participant grows weary may lead a participant to get careless, skip items, or leave the assessment altogether (McColl, Eccles, Rousseau, Steen, Parkin, & Grimshaw, 2003, p. 778). The decision to have consistent ordering placement of the six anchor items was informed by equating recommendations found in test construction literature (Brennan & Kolen, 1987).

As much as possible, the types of questions were matched on both forms. For example, if an N predictive item was placed as item 8 on Form BC, a similar predictive item was placed as item 8 on Form LC. The decision not to randomize the delivery of the
items within the form itself was based on current literature related to the potential impact of item placement on the difficulty and discrimination measures (Fowler, 1995; Impara & Foster, 2006).

Table 3

Breakdown of the 54 item pool into 2 forms

<table>
<thead>
<tr>
<th>Form</th>
<th>&quot;N&quot; unique items</th>
<th>&quot;A&quot; unique items</th>
<th>&quot;N &amp; A&quot; anchor items</th>
<th>Total items</th>
</tr>
</thead>
<tbody>
<tr>
<td>Big C</td>
<td>13</td>
<td>11</td>
<td>4 &amp; 2</td>
<td>30</td>
</tr>
<tr>
<td>Little c</td>
<td>13</td>
<td>11</td>
<td>4 &amp; 2</td>
<td>30</td>
</tr>
</tbody>
</table>

The elcapstone.com website link went active on September 1, 2010 and throughout the month anonymous volunteers were able to participate in the EL-Capstone online administration of the Stage 2 Participant scale development data collection phase. The recruitment effort was supported through various social networks to reach various field-related resources and anonymous volunteers to cover a wide range of possible early literacy knowledge levels. Appendix D has examples of recruitment postings that supported the online administration of EL-Capstone.

The elcapstone.com script counter directed the delivery of the forms; however, it was the SelctSurvey.ASP software counters that captured the number of times a survey was started. Unfortunately, a more accurate internal counter was not available in the software to count the number of surveys that were actually completed.

4.3 Stage Three: Post-participant

This post-participant stage centered on conducting an analysis of the results from the online administration of EL-Capstone. The information extracted from this analysis phase is used to inform decisions regarding the selection and reduction of the instrument items that are to move forward into the next EL-Capstone research project.

The EL-Capstone online administration response rate allowed for either the item response theory or the classical test theory to serve as the theoretical framework. The
differences of CTT and IRT were discussed earlier in the literature review of this study, i.e., 2.3 Theoretical Considerations, however, this section will highlight a few of these characteristics as they relate to the analysis processes of this study. The IRT definitional regression function focus on the unidimensional latent trait, i.e., \( \theta \), made IRT the stronger fit over CTT. One strength associated with using IRT is a posture that the associated latent ability (\( \theta \)) is not bound to the unique cluster of test items contained within the specific test form, and consequently, test items may be added or dropped without generating statistical errors. These model characteristics were influential determinants in the selection of IRT for the analysis phase of this study.

4.3.1 Item Response Theory

IRT, a latent trait model, has been widely used since the 1960s. Unlike CTT that is sample dependent, the IRT models are more invariant over population samples (Hambleton, 2004). The latent trait in this study is the content knowledge ability of early literacy concepts (EL). This ability is not directly observable and is being measured indirectly through this multi-item scale. The underlying latent trait is represented by the Greek letter Theta (\( \theta \)). The person's ability is being modeled in probabilistic terms and it is this latent trait alone that is assumed to account for the variances in the responses in this scale. More specifically, this assumption is known as the unidimensionality assumption. EL-Capstone, for this study, was being designed to develop scale items that would capture the existing early literacy core knowledge concepts levels of adults.

One strength associated with using IRT is a postulate that the associated latent ability (\( \theta \)) is not bound to the unique cluster of test items contained within the specific test form, and consequently, test items may be added or dropped without generating statistical errors. As a larger number of items were included in this initial scale development phase, the characteristics of IRT made it a strong model for use with EL-Capstone. There was no reason to presume that there were any major differences between these two anonymous adult groups; however, IRT extends CTT limitations, especially when used in conjunction with equating methods.

IRT has associated assumptions. It assumes unidimensionality and local independence. As true-false items, i.e., dichotomously scored items, correct answers received a score of one. EL-Capstone was designed to measure a single latent ability, i.e.,
emergent literacy core knowledge, and this one trait determined the success on each item. IRT is focused on whether or not an individual item is correctly answered and not on the participant's total score; therefore, one may say that the focus is on how the individual items perform and a participant's probability of correctly answering each item. Baker illustrates this point by stating "the probability of a correct response is near zero at the lowest levels of ability" (2001, p. 7).

As shown in Figure 15 based on the Baker (2001) description, EL-Capstone took a response-centered approach to IRT. In discussing the concept of scales, Baker highlights "the theoretical range of ability is from negative infinity to positive infinity, practical considerations usually limit the range of values from, say, -3 to +3 (2001, p. 6). The -3 to +3 scale was used in this study for the x-axis.

![Figure 15](image)

*Figure 15. A scaling diagram demonstrating a response-centered approach.*

This scaling method attempts to place a person's ability level ($\theta_i$) and the difficulty ($b_j$) on the same scale. The notion that matched ability groups should have an equal probability of correctly responding to an item is central to this scale. In discussing the benefits of IRT during scale construction, Molenaar (1997) highlights the ability to identify items that need revisions or may need to be removed from the scale altogether.

In an IRT 1-PL model, only the $b$ parameter, or difficulty, is allowed to vary. This translates into an assumption that all items have equivalent discrimination. Both the discrimination parameter ($a$) and the difficulty vary in a 2-parameter model (2-PL) while the pseudo-guessing parameter ($c$) is held constant; therefore, $c=0$. A 3-PL model is
appropriate for large dichotomous data sets where guessing may have a significant impact. The 2-PL model may be mathematically represented as:

\[ P(y_j = 1 | \theta) = \frac{e^{D_{ij}(\theta - b_j)}}{1 + e^{D_{ij}(\theta - b_j)}} \]

The IRT definitional regression function focus on the unidimensional latent trait, i.e., \( \theta \), made IRT the stronger fit over CTT; therefore, a 2-PL model was used in this study and the pseudo-guessing parameter was held constant.

4.3.2 Data Preparation

The output reports for the SurveySelect.Asp software were downloaded and backups of the data were archived to a secure external hard drive. The user response report format for the two forms, i.e., Big C: EL-Capstone (BC) and Little c: EL-capstone (LC), were saved as Microsoft Office Excel comma separated values files. Upon first inspection of Form LC, an internal if/then auto-generated coding error by the SurveySelect.Asp software was discovered. Hidden fields can be added only before a survey has had any responses. The administrative function of the software indicated that no hidden fields had been configured for either of the forms. However, the internal error caused the first item in Form LC to be skipped when the participant responded to an optional preceding demographic item. A decision was made to drop the item from the analysis phase since only five responses had been generated for this item.

The data in both forms, BC and LC, were examined visually for patterns related to missing response data. Missing data holds information. For example, in addition to showing overall participant attrition, items with unusually high number of missing responses may indicate a particular difficulty was encountered in comprehending the content of that particular item. The visual pattern search was completed within Excel through a find and replace of the "No Answer" values with blank cells. Reducing the viewing magnification to 20% created an opportunity to easily observe patterns across all records. Not finding any obvious patterns related to missing data, a conservative approach was adopted to drop all records that contained "not answered" or missing data.

The two Excel reports were each imported into the statistical analysis software, SAS 9.1, where all of the incomplete records were purged. The SurveySelect.Asp user
response report only generated a TRUE/FALSE/No Answer output and the report did not indicate the correctness value associated with that response. Therefore, SAS was used to adjust the correctness values for the "False" items.

Using a statistical modeling program, Mplus 3.01, a 2-PL summary of analysis for Form BC was conducted on one group with 443 observations, 30 dependent variables, zero independent variables, and one continuous latent variable. A 2-PL summary of analysis for Form LC was conducted on one group with 381 observations, 29 dependent variables, zero independent variables, and one continuous latent variable.

4.3.3 Item Parameters

The $a$ parameter is a slope of a logistic curve. It is used to predict the probability of a person to correctly answer an item based on $\theta$, i.e., ability. Since correctly answering assumes a positive relationship with $\theta$, positive values were expected. Subsequently, negative discriminations are not good. This indicates that examinees of higher ability have a lower probability of correctly responding to an item as compared to examinees of lower ability. Generally these items should be discarded (or perhaps completely revised) before additional administrations.

The $b$ parameter represents the difficulty level of the item. This parameter is also commonly referred to as the threshold parameter (Reeve & Mâsse, 2004) or the difficulty index (Baker, 2001) because it describes the location of theta ($\theta$) along the x-axis. There are subtle differences in these terms; however, they are used to discuss the relationship of ability to performance probability. For example, the term threshold is the probability of getting the item wrong when $\theta$ is equal to zero. The $b$ parameter is the amount of $\theta$ one needs at the point where the line crosses 0.5. While negative difficulties are acceptable, generally extremely high difficulties (positive or negative) are usually bad. Appendix E contains the two parameters, $a$ and $b$, for both BC and LC forms.

Instead of using raw total scores of the thresholds for the difficulty parameter, real values of each item difficulty on an IRT scale were calculated by the difficulty parameter divided by the discrimination. For example, the formula for the real value of difficulty on an IRT scale for item N47 ($b = -1.504$) was calculated using $2.906/1.598$. Appendix E contains the IRT adjusted difficulty parameters for both BC and LC forms. This adjustment process was required since the MPlus 3.01 software access version at this
time did not automatically calculate the IRT parameterization in two-parameter logistic metric where the logit is $1.7\times$discrimination*(theta - difficulty). A logit is a natural log of the odds ratio. As a note, IRT does not operate directly on probability but rather, it gets transformed into something more continuous by using either logit or probit (Lee, Song, & Cai, 2010; Stocking & Lord, 1983).

The means of the $a$ and $b$ parameters and the difficulty standard deviation (STD) were calculated for both forms. The means for Form BC were $a = 0.568$ and $b = -3.109$ and the STD $B = 4.340$. The means for Form LC were $a = 0.129$ and $b = -0.237$ and the STD $B = 32.050$. Two forms and two groups on different scales made comparisons of the items difficult. Equating made comparisons possible on the same scale.

**4.3.4 Equating**

A relationship between inventories that differ in difficulty or content may be established through linking methods, e.g., equating (Kim, 2007). Item response theory allows for scores from one form to be equated to the scores on another form (Cook & Eignor, 1991; Hambleton & Swaminathan, 1985). In classical equating methods, this would mean that the two EL-Capstone forms administered online would be scaled so regardless of the form the participant completed, that person's same ability estimate could translate to the items on the other form that was not completed. In IRT equating designs, the common linking items place item parameter estimates on the same scale (Cook & Eignor, 1991). Mislevy in discussing IRT linking and equating when the parameters are known posits "the varying degrees of difficulty and accuracy among test forms are accounted for by the different parameters of the items that comprise them (1993, p. 57)."

There are three general designs recommended for use in IRT equating; (a) single group, (b)random groups, and (c) anchor test (Cook and Eignor, 1991). Since there were multiple groups and they were not randomly assigned, the anchor test linking design was used. Cook and Eignor posit that when using an anchor test design the groups "do not need to be randomly equivalent, and usually they are not" (1991, p. 193). For this study, two groups of volunteer participants took one of the two forms of the instrument. Both forms contained 24 unique items and a set of six internal anchor items. Scale linking has obvious benefits of increasing the participant sample size and total number of items (Chen, Revicki, Lai, Cook, & Amtmann, 2009 ).
In a discussion of some practical issues in equating, Brennan and Kolen suggest that "equating is improved when the common items linking these forms constitute a miniature version of the full-length test" (1987, p. 284). EL-Capstone's six linking items consisted of both "N" statements and "A" statements and represented the same True/False format of the non-linking items in each form. As defined by Cook and Eignor, the six anchors used in the 30 item online EL-Capstone administration task met the generally accepted standard of being "at least as long as 20% of the total test length" (1991, p. 193).

Spiraling of the two instrument forms were completed online by anonymous adult participants; however, the participants were still treated as nonequivalent groups; therefore, equating places the parameters from the two forms on the same scale through a linear transformation (Cook & Eignor, 1991; Dorans, 1990). Baker and Al-Karni recommend that the "initial task in IRT test equating is to determine the slope and the intercept of the linear equation used to transform the metric of one calibration to that of another" (1991, p. 148).

The six linking item discrimination and difficulty parameters, as seen in Appendix F, determined the transformation coefficients used to convert form Little c into the scale of form BC. The transformation coefficients were $a = 0.135$ and $b = -3.077$. These were calculated by plugging Form BC's mean $b$ (-3.109) and STD B (4.340) with Form LC's mean $b$ (-0.129) and STD B (32.050) into the equating test scores formula as described by Cook and Eignor (1991, p. 195)

$$\hat{T}_x = \sum_{i=1}^{n_x} \hat{P}_i(\theta), \quad (8)$$

$$\hat{T}_y = \sum_{i=1}^{n_y} \hat{P}_j(\theta), \quad (9)$$

where

$\hat{T}_x = \text{Test form X estimated true score},$

$\hat{T}_y = \text{Test form Y estimated true score, and } \hat{P}_i(\theta) \text{ and } \hat{P}_j(\theta) \text{ are the estimated item-response functions for items } i \text{ (in test form X) and } j \text{ (in test form Y), respectively.}$

According to Cook and Eignor (1991):

Equations (8) and (9) can be used to transform any $\theta$ (not just $\theta$s actually obtained from the administration of the test forms) to an estimated true score on the respective test forms. Because of this and because the item parameter estimates used in (8) and (9) are
independent of the groups taking the tests, the conversion, or relationship between the true scores on the two forms, can also be said to be independent of the groups used to obtain it. (p. 195)

Using the Cook and Eignor equations, the EL-Capstone Forms BC and LC \( \theta \)s were transformed to estimated true scores as follows:

\[
\frac{C_1 - (-3.1087)}{4.329694} = \frac{C_2 - (-0.23701)}{32.05003}
\]

\[
C_1 = \frac{4.329694}{32.05003} (C_2 = 0.23701) - 3.1087
\]

\[
= 0.1354037422 (C_2 + 0.23701) - 3.1087
= 0.1354 C_2 + 0.032092 - 3.1087
= 0.1354 C_2 - 3.07660
\]

After the estimated true scores relationship of Form BC and LC was established, all of the LC parameters were transformed. For example, Form LC Item N6 parameter \( a \) (0.375) transformed into Form BC by calculating \( \frac{0.375}{0.135} = 2.770 \). The \( b \) parameter (-4.877) was transformed by the equation \( 0.135 \times -4.877 - 0.237 = -0.897 \). All of the transformations of Form LC into the scale of Form BC are shown in Appendix G.

4.3.5 Item Characteristic Curves (ICC)

In an ICC curve, the latent trait is represented along the horizontal x-axis by logistic curves that model the probability of a person correctly responding to an item in this scale. Individuals with low EL levels are to the left of this axis. Moving to the right across the axis identifies people that have higher EL knowledge levels as identified by the item responses in this scale. So in IRT a person's Theta or ability is estimated and under the assumptions of unidimensionality, this factor alone is said to account for the variance in responses. Baker posits the "item characteristic curve is the basic building block of item response theory; all the other constructs of the theory depend upon this curve" (2001, p. 7). This mathematical expression linked the participant's probability of an expected score on the latent trait, i.e., early literacy core knowledge.

The ICC is a visual representation of the item's performance regressed on the participant's ability. Based on the difficulty parameter, if an item is easy, people with low \( \theta \)s, or low ability, are expected to correctly answer an item. On the other end of the
difficulty location index, difficult items are expected to be correctly answered by people with high ability, or high $\theta$s. The ICC drawing in Figure 16 is provided to clarify the concept of the relationship of an attribute to the likelihood of endorsement. IRT is a non-linear model bound at both ends and creates a characteristic s-shaped slope.

![Figure 16](image)

*Figure 16.* An ICC showing the relationship of an attribute to the likelihood of endorsement.

Figure 17 illustrates an ICC for an easy performing item in EL-Capstone. Across ability, this item demonstrated a high likelihood of a correct response. With little change in the probability of a person correctly answering Item N34, this item would not move forward without revision to the next administration of the instrument.

![Figure 17](image)

*Figure 17.* An ICC example of a low difficulty item.
The steepness of the ICC discriminates how well an item is able to highlight differences in the latent abilities of the participants. The steeper the curve, the greater the item's ability to discriminate. As seen in Figure 17, Item N34 is easy and it is not discriminating well across those with high and low early literacy core knowledge abilities as measured by EL-Capstone. Figure 18 presents an item that discriminated well across abilities.

![Image of Item Characteristic Curve](image)

**Figure 18.** An ICC example of an item discriminating well across ability (θ).

Item characteristic curves (ICC) were constructed for each of item. All ICC information from the pre-reversed coded results are available in Appendix H. The curves are displayed in conjunction with the 2-PL information for all items.

### 4.3.6 Unidimensionality Investigation

IRT is designed to account for the covariance between items. Each item has a mean and gets its own intercept (slope). Each item has variance and its own error variance. IRT allows for the examination of all possible covariance between items. One of the assumptions in IRT is unidimensionality. This assumption means that the covariance among the items is based on the expected factor. In this study, it was assumed that theta, level of early literacy knowledge, was the factor explicitly accounting for the differences in discrimination and difficulty. Item construction was bound by the scope of
 Unlike the assumptions of CTT, IRT allows for testing of the assumptions (Segars, 1997).

An exploration of the structure of the construct being measured, i.e., the latent construct of early literacy knowledge factor structure was conducted through an exploratory factor analysis (EFA) run in Mplus (version 4.21). The EFA was not used to confirm the structure but to support the assumption of this unidimensionality. Muthén and Muthén state, "Exploratory factor analysis (EFA) is used to determine the number of continuous latent variables that are needed to explain the correlations among a set of observed variables (1998-2010, p.41). EFA output files for both the BC and the LC forms suggested there were multiple factors. For the percent of variance accounted for by the first factor, the first eigenvalue from the EFA output was divided by the number of items in the form. A resultant value of greater than 40% is considered as acceptable unless the second factor has a high eigenvalue (Zickar, Russell, Smith, Bohle, & Tilley, 2002). The resultant values from the first eigenvalues did not meet these levels, i.e., BC had 13% and LC had 17%.

While it could still be argued that the items were bound by the scope of the NELP Report (2008a), alternative investigations included looking at the varimax rotated loadings output from the EFA number of factors with a chi-square probability value at least with a .01 significance from each of the two EL-Capstone forms. It should be noted that there is some discussion in the literature that dichotomous scale components must be parallel (tau-equivalent) for the coefficient alpha to be a dependable estimator of scale reliability (Raykov, Dimitrov, & Asparouhov, 2010). Clustering the items and then examining the items for possible patterns related to the item types was not supported by the varimax rotated loadings. For example, there were items that were predominately definitional, predictive, or activity-based; however, clustering on these areas was not supported by the loadings.

Hypothetically, the nature of the early literacy domain itself could be argued to be multiunidimensional since there are known sub-domains, e.g., phonological awareness or alphabetic knowledge. Some researchers argue a case for building a unidimensional test using multidimensional items (Reckase, Ackerman, & Carlson, 1988; Sheng & Whikle, 2007). Discussing alternatives for when unidimensionality assumptions are not met,
Nandakumar suggest "one can perhaps turn to the theory of testlet scoring (Thissen, Steinberg, & Mooney, 1989), which is theoretically more appropriate; use the Reckase (1985, 1989) multidimensional compensatory logistic modeling approach; or break up the test into several essentially unidimensional subtests, each scored separately" (1991, p.11).

In addition to investigating the clustering of items, psychometric literature on scale item development was consulted. Spector, VanKatwyk, Brannick, and Chen (1997) note that the production of two factors can sometimes arise from the inclusion of both negative and positively worded items. The influence of careless responding to reverse-worded items during factor analysis is also discussed by Woods (2006), and the author posits that these items may cause researchers to falsely reject the unidimensional assumption of IRT models. Forms BC and LC contained negated items. Four items were re-coded from the original data files that included the missing data. The new data files for forms BC and LC were re-named BCmr and LCmr, respectively.

An EFA was run on BCmr and LCmr. The resultant values from the first eigenvalues exceeded the 40% standard (Sinar & Zickar, 2002). BCmr's 60% value was calculated by taking the first eigenvalue of 17.9 and dividing it by 30 items. Figure 19 displays the eigenvalue plot of BCmr.

Figure 19. BCmr eigenvalue plot.
LCmr’s 64% value was calculated by taking the first eigenvalue of 18.521 and dividing it by 29 items. Figure 20 illustrates the LCmr eigenvalue plot. These exploratory analyses indicated sufficient levels of unidimensionality of both BCrr and LCmr scales for an IRT analysis.

![LCmr eigenvalue plot](image)

*Figure 20. LCmr eigenvalue plot*

Once the unidimensional assumption of IRT was met, an IRT analysis was run in Mplus (version 4.21) using the reversed coded data, i.e., BCmr and LCmr. It should be noted that unlike the earlier version used in this study, Mplus version 4.21 automatically computed the logistic model and output the item discriminations and item difficulties. Appendix I displays the ICCs for the reversed coded BCmr and LCmr.

### 4.3.8 Reverse Coded Item Discussion & Summary

An IRT analysis was run in Mplus (version 4.21) using the reversed coded data, i.e., BCmr and LCmr. When using Mplus to conduct a two-parameter logistic item response theory (IRT) model, the categorical data output may be set to include the item discriminations and item difficulties. The discriminations and item difficulty results from Form BCmr and LCmr are shown in Table 4 and Table 5, respectively.
The discrimination parameter describes the strength of the scale item to differentiate among people at different levels of Theta. As a result of the response data collected from n=443 adults, it was found that Form BCmr had 19 items (N16, A1, N21, A23, A25, N42, N50, N51, N34, A20, N48, A26, A8, A30, A33, A27, N33, N4, N5) with unacceptable item discriminations, \( a \) parameters. Unacceptable is defined as levels of \(<0.3 \) (Ryu, Suh, & Kim (2009) while typical acceptable range levels are between 0.5 and 2.5 (Reeve & Fayers, 2005).

For this first round of scale development, acceptable levels in this study are defined as \( \geq 0.3 \) and \( \leq 2.5 \). This range was selected to minimize items being removed too early in the developmental process without additional testing. Eleven items met the acceptable range (N1, A31, N41, A16, A28, N11, N44, A5, N52, N14, and N47). This means these items are differentiating among people of different levels of EL knowledge, i.e., Theta.

Negative discriminations are bad thing in IRT because this means that examinees of higher ability have a lower probability of correctly responding to an item as compared to examinees of lower ability. Generally these items are discarded or perhaps completely revised. Form BCmr had 9 items (N16, A1, N21, A23, A25, N42, N50, N51, N34) with negative discriminations that ranged from -0.478 to -0.021.

One explanation for the source of these negative discriminations may stem from individuals with higher EL ability that may have knowledge that creates an "it depends" response. For example, item N51 has a -0.028 discrimination parameter. The N51 item statement, There is research-based evidence that storytelling is a strong predictor in later reading comprehension, may be eliciting interpretations related to the definition of research-based evidence, storytelling, or even later reading comprehension. While the causes of these negative discriminations are unknown at this time, these nine items will not be moved to the next administration without revisions.

As discussed by Embretson and Reise, "item difficulty becomes more extreme as item discriminations decrease" (2000, p. 70). Recall that the difficulty (\( b \)) parameter in a 2-PL ICC chart indicates the location of where an item's difficulty is relative to the ability. For example, if an item's difficulty equals a person's ability (\( \theta \)), then that person has a 50/50 chance of getting the item correct. Therefore, if an item has a difficulty
greater than a person’s ability, they are not likely to get the item correct. The items are ordered by difficulty. When there are multiple items with the same difficulty level, the item with the steepest slope (or highest discrimination) is kept. Since all of the items are measuring the same ability, the actual wording of the item is secondary to the mathematical modeling outcomes produced through an IRT analysis.

Table 4

*Mplus discrimination (a) and difficulty (b) parameters from the Form BCmr data along with the number of item responses and the percentage correct.*

<table>
<thead>
<tr>
<th>BCmr Item</th>
<th>a</th>
<th>b</th>
<th>Total Number of Respondents</th>
<th>Percentage Correct</th>
</tr>
</thead>
<tbody>
<tr>
<td>A31</td>
<td>0.334</td>
<td>-3.134</td>
<td>593</td>
<td>84</td>
</tr>
<tr>
<td>N47</td>
<td>0.938</td>
<td>-1.506</td>
<td>578</td>
<td>84</td>
</tr>
<tr>
<td>A20</td>
<td>0.034</td>
<td>49.904</td>
<td>573</td>
<td>5</td>
</tr>
<tr>
<td>N5</td>
<td>0.293</td>
<td>-4.614</td>
<td>563</td>
<td>90</td>
</tr>
<tr>
<td>N4</td>
<td>0.278</td>
<td>-4.644</td>
<td>556</td>
<td>89</td>
</tr>
<tr>
<td>A30</td>
<td>0.192</td>
<td>-7.754</td>
<td>551</td>
<td>91</td>
</tr>
<tr>
<td>N11</td>
<td>0.522</td>
<td>-2.781</td>
<td>544</td>
<td>90</td>
</tr>
<tr>
<td>A27</td>
<td>0.246</td>
<td>-5.817</td>
<td>545</td>
<td>91</td>
</tr>
<tr>
<td>N41</td>
<td>0.414</td>
<td>-4.494</td>
<td>535</td>
<td>95</td>
</tr>
<tr>
<td>N1</td>
<td>0.323</td>
<td>-5.208</td>
<td>531</td>
<td>95</td>
</tr>
<tr>
<td>A5</td>
<td>0.65</td>
<td>-0.894</td>
<td>515</td>
<td>70</td>
</tr>
<tr>
<td>N14</td>
<td>0.911</td>
<td>-1.766</td>
<td>511</td>
<td>88</td>
</tr>
<tr>
<td>A26</td>
<td>0.162</td>
<td>-11.192</td>
<td>512</td>
<td>95</td>
</tr>
<tr>
<td>N33</td>
<td>0.251</td>
<td>-2.698</td>
<td>508</td>
<td>75</td>
</tr>
<tr>
<td>A28</td>
<td>0.459</td>
<td>-1.449</td>
<td>506</td>
<td>74</td>
</tr>
<tr>
<td>N34</td>
<td>-0.021</td>
<td>23.31</td>
<td>505</td>
<td>70</td>
</tr>
<tr>
<td>A16</td>
<td>0.428</td>
<td>-4.937</td>
<td>504</td>
<td>97</td>
</tr>
<tr>
<td>N44</td>
<td>0.596</td>
<td>-1.902</td>
<td>506</td>
<td>84</td>
</tr>
<tr>
<td>N52</td>
<td>0.757</td>
<td>-1.186</td>
<td>505</td>
<td>76</td>
</tr>
<tr>
<td>A33</td>
<td>0.221</td>
<td>-7.534</td>
<td>499</td>
<td>94</td>
</tr>
<tr>
<td>N50</td>
<td>-0.079</td>
<td>-8.543</td>
<td>503</td>
<td>77</td>
</tr>
<tr>
<td>A1</td>
<td>-0.455</td>
<td>1.064</td>
<td>554</td>
<td>67</td>
</tr>
<tr>
<td>N48</td>
<td>0.154</td>
<td>-4.548</td>
<td>536</td>
<td>76</td>
</tr>
<tr>
<td>A25</td>
<td>-0.118</td>
<td>-3.163</td>
<td>535</td>
<td>33</td>
</tr>
<tr>
<td>N16</td>
<td>-0.478</td>
<td>-0.833</td>
<td>524</td>
<td>35</td>
</tr>
<tr>
<td>N21</td>
<td>-0.255</td>
<td>-2.062</td>
<td>507</td>
<td>29</td>
</tr>
<tr>
<td>N42</td>
<td>-0.111</td>
<td>7.644</td>
<td>514</td>
<td>81</td>
</tr>
<tr>
<td>A8</td>
<td>0.188</td>
<td>-7.804</td>
<td>513</td>
<td>92</td>
</tr>
<tr>
<td>A23</td>
<td>-0.239</td>
<td>3.811</td>
<td>507</td>
<td>81</td>
</tr>
<tr>
<td>N51</td>
<td>-0.028</td>
<td>-37.578</td>
<td>503</td>
<td>14</td>
</tr>
</tbody>
</table>
Form BCmr had 25 items (N51, A26, N50, A8, A30, A33, A27, N1, A16, N4, N5, N48, N41, A25, A31, N11, N33, N21, N44, N14, N47, A28, N52, A5, N16) with negative difficulty levels. These negative values ranged from -0.833 to -37.578. While negative difficulties are acceptable, it does indicate that the items are being perceived as an easier item to correctly answer. Four items (N51, A26, N34, A20) had extremely high difficulties (positive or negative) that were double digits (-37.578, -11.192, 23.31, 49.904). These extremes are considered problematic as the individuals tend to stack up at the extreme locations along the axis. These four extreme \( b \) parameters were examined for possible miscoding of the correctness value; however, they were found to be correctly coded.

Table 5 presents the results of the Mplus output from the response data collected from \( n=381 \) adults participating in the study through Form LCmr. The Mplus analysis indicates that Form LCmr had 18 items (NN49, N15, AA17, NN12, N21, N42, N36, AA14, N16, A32, A22, NN17, A21, N37, N2, N6, A5, A15) with unacceptable item discriminations, \( a \) parameters. Eleven items fell within the acceptable (\( \geq 0.3 \) and \( \leq 2.5 \)) discrimination range (N10, A24, N1, A8, A7, N43, NN46, A29, A11, N28, and N23). These 11 items were evaluated relative to their difficulty parameters.

As mentioned earlier, negative discriminations are bad in IRT because this means that examinees of higher ability have a lower probability of correctly responding to an item as compared to examinees of lower ability. Form LCmr had 7 negative discrimination items (NN49, N15, AA17, NN12, N21, N42, and N36) that ranged from -0.082 to -0.861. Six of the seven Form LCmr negative discrimination items were of type \( N \). You may recall from an earlier discussion related to the construction of the EL-Capstone items that the \( N \) items have a stronger direct connection to the wording of the NELP Report than the \( A \) items. Item N36 with a -0.082 is an example of a negative discrimination item from Form LCmr. The N36 wording is a direct connection to the NELP 2008 Report text, i.e., "Complex forms of oral language skills (e.g., grammar and definitional vocabulary) were determined to be better predictors of later literacy than were simple vocabulary" (2008, p.193). The source of these negative discriminations is not known; however, these 7 items will not be moved to the next administration without revisions to the text or format.
Table 5

*Mplus* discrimination (*a*) and difficulty (*b*) parameters from the Form LCmr data along with the number of item responses and the percentage correct.

<table>
<thead>
<tr>
<th>LCmr Item</th>
<th>a</th>
<th>b</th>
<th>Total Number of Respondents</th>
<th>Percentage Correct</th>
</tr>
</thead>
<tbody>
<tr>
<td>N2</td>
<td>0.182</td>
<td>-4.868</td>
<td>520</td>
<td>81</td>
</tr>
<tr>
<td>A32</td>
<td>0.04</td>
<td>-7.78</td>
<td>501</td>
<td>62</td>
</tr>
<tr>
<td>N6</td>
<td>0.221</td>
<td>-4.872</td>
<td>477</td>
<td>86</td>
</tr>
<tr>
<td>A21</td>
<td>0.149</td>
<td>-3.623</td>
<td>475</td>
<td>69</td>
</tr>
<tr>
<td>N10</td>
<td>0.341</td>
<td>-2.037</td>
<td>472</td>
<td>74</td>
</tr>
<tr>
<td>A7</td>
<td>0.424</td>
<td>-4.323</td>
<td>467</td>
<td>95</td>
</tr>
<tr>
<td>N15</td>
<td>-0.611</td>
<td>-1.61</td>
<td>457</td>
<td>20</td>
</tr>
<tr>
<td>A22</td>
<td>0.041</td>
<td>-9.37</td>
<td>457</td>
<td>67</td>
</tr>
<tr>
<td>N36</td>
<td>-0.082</td>
<td>2.48</td>
<td>441</td>
<td>57</td>
</tr>
<tr>
<td>N1</td>
<td>0.383</td>
<td>-3.874</td>
<td>448</td>
<td>91</td>
</tr>
<tr>
<td>A5</td>
<td>0.249</td>
<td>-2.442</td>
<td>435</td>
<td>73</td>
</tr>
<tr>
<td>A24</td>
<td>0.376</td>
<td>-3.735</td>
<td>431</td>
<td>90</td>
</tr>
<tr>
<td>N28</td>
<td>2.54</td>
<td>-2.102</td>
<td>435</td>
<td>98</td>
</tr>
<tr>
<td>A15</td>
<td>0.289</td>
<td>-6.215</td>
<td>432</td>
<td>95</td>
</tr>
<tr>
<td>N37</td>
<td>0.178</td>
<td>-1.162</td>
<td>431</td>
<td>41</td>
</tr>
<tr>
<td>A11</td>
<td>1.401</td>
<td>-2.712</td>
<td>432</td>
<td>99</td>
</tr>
<tr>
<td>N43</td>
<td>0.468</td>
<td>-4.366</td>
<td>430</td>
<td>95</td>
</tr>
<tr>
<td>A29</td>
<td>0.662</td>
<td>-3.035</td>
<td>431</td>
<td>95</td>
</tr>
<tr>
<td>N23</td>
<td>3.095</td>
<td>-2.154</td>
<td>431</td>
<td>98</td>
</tr>
<tr>
<td>AA17</td>
<td>-0.33</td>
<td>-2.965</td>
<td>518</td>
<td>16</td>
</tr>
<tr>
<td>NN46</td>
<td>0.499</td>
<td>-4.421</td>
<td>512</td>
<td>96</td>
</tr>
<tr>
<td>NN12</td>
<td>-0.326</td>
<td>-1.569</td>
<td>443</td>
<td>27</td>
</tr>
<tr>
<td>AA14</td>
<td>0.013</td>
<td>56.487</td>
<td>433</td>
<td>72</td>
</tr>
<tr>
<td>NN17</td>
<td>0.105</td>
<td>-7.356</td>
<td>430</td>
<td>90</td>
</tr>
<tr>
<td>N16</td>
<td>0.031</td>
<td>-42.779</td>
<td>440</td>
<td>30</td>
</tr>
<tr>
<td>N21</td>
<td>-0.317</td>
<td>-1.781</td>
<td>433</td>
<td>23</td>
</tr>
<tr>
<td>N42</td>
<td>-0.103</td>
<td>-5.534</td>
<td>435</td>
<td>79</td>
</tr>
<tr>
<td>A8</td>
<td>0.388</td>
<td>-3.542</td>
<td>436</td>
<td>90</td>
</tr>
<tr>
<td>NN49</td>
<td>-0.861</td>
<td>-2.843</td>
<td>429</td>
<td>3</td>
</tr>
</tbody>
</table>

Form LCmr had 27 items (N16, A22, A32, NN17, A15, N42, N6, N2, NN46, N43, A7, N1, A24, A21, A8, A29, AA17, NN49, A11, A5, N23, N28, N10, N21, N15, NN12, N37) with negative difficulty levels. These negative values ranged from -1.162 to -42.779. Again, negative difficulties are acceptable; however, it does indicate that the items are being perceived as an easier item to correctly answer. Item N16 had an
extremely high negative $b$ parameter (-42.779) and AA14 had a double digit positive value of 56.487. These two extreme $b$ parameters were examined for possible miscoding of the correctness values; however, they were found to be correctly coded.

As discussed by Hambleton (1989), the IRT model creates an estimate of an individual's ability that is independent of the items. Cook and Eignor write "... it does not matter if an examinee takes an easy or hard form of the test, the examinee's ability estimates will be identical, within measurement error, provided the parameter estimates for both forms have been placed on the same scale" (1991, p. 192). As in the original processing, the reversed coded data needed to be transformed to have the items on both forms placed on the same scale. The six linking item discrimination and difficulty parameters, as seen in Table 6, determined the transformation coefficients used to convert form LCmr into the scale of form BCmr.

Table 6

*The six common linking item information used to establish transformation coefficients to equate LCmr into the scale of BCmr.*

<table>
<thead>
<tr>
<th>Item</th>
<th>BCmr</th>
<th>LCmr</th>
</tr>
</thead>
<tbody>
<tr>
<td>N1</td>
<td>0.323</td>
<td>-5.208</td>
</tr>
<tr>
<td>A5</td>
<td>0.065</td>
<td>-0.894</td>
</tr>
<tr>
<td>N16</td>
<td>-0.478</td>
<td>-0.833</td>
</tr>
<tr>
<td>N21</td>
<td>-0.255</td>
<td>-2.062</td>
</tr>
<tr>
<td>N42</td>
<td>-0.111</td>
<td>7.644</td>
</tr>
<tr>
<td>A8</td>
<td>0.188</td>
<td>-7.804</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>BCmr</th>
<th>LCmr</th>
</tr>
</thead>
<tbody>
<tr>
<td>mean a</td>
<td>-0.04467</td>
<td>0.105167</td>
</tr>
<tr>
<td>mean b</td>
<td>-1.52617</td>
<td>-9.992</td>
</tr>
<tr>
<td>std b</td>
<td>5.257876</td>
<td>16.11407</td>
</tr>
</tbody>
</table>

Using the Cook and Eignor (1991) IRT equating method, the reverse coded Forms BC (BCmr) and LC (LCmr) thetas were transformed to estimated true scores as follows:

$$\frac{c_1 - (-1.52617)}{5.257876} = \frac{c_2 - (-9.992)}{16.11407}$$
\[
C_1 = \frac{5.257876}{16.11407} (C_2 = 9.992) - 1.52617 \\
= 0.326291 (C_2 + 9.992) - 1.52617 \\
= 0.3263 C_2 + 3.26030 - 1.52617 \\
= 0.3263 C_2 + 1.73413
\]

Consistently throughout this study, Form LCmr items are transformed into Form BCmr. After the estimated true scores relationship of BCmr and LCmr were established, all of the LCmr parameters were transformed. Using the same item example from the non reversed data, LCmr's Item N6 discrimination parameter (0.221) transformed into BCmr through the following calculation: \(0.221 / 0.3263 = 0.677291\). The difficulty parameter (-4.872) was transformed by the equation \(0.3263 \times -4.872 - 9.992 = -11.5817\).

The transformed discrimination \((a)\) and difficulty \((b)\) parameters on the reversed coded data for LCmr are shown in Table 7.

**Table 7**

*Shown are the new discrimination \((a)\) and difficulty \((b)\) parameters on the reversed coded data for LCmr with the linking items noted by shaded blank records.*

<table>
<thead>
<tr>
<th>Item</th>
<th>(a)</th>
<th>(b)</th>
</tr>
</thead>
<tbody>
<tr>
<td>N2</td>
<td>0.558</td>
<td>-11.580</td>
</tr>
<tr>
<td>A32</td>
<td>0.123</td>
<td>-12.531</td>
</tr>
<tr>
<td>N6</td>
<td>0.677</td>
<td>-11.582</td>
</tr>
<tr>
<td>A21</td>
<td>0.457</td>
<td>-11.174</td>
</tr>
<tr>
<td>N10</td>
<td>1.045</td>
<td>-10.657</td>
</tr>
<tr>
<td>A7</td>
<td>1.299</td>
<td>-11.403</td>
</tr>
<tr>
<td>N15</td>
<td>-1.873</td>
<td>-10.517</td>
</tr>
<tr>
<td>A22</td>
<td>0.126</td>
<td>-13.049</td>
</tr>
<tr>
<td>N36</td>
<td>-0.251</td>
<td>-9.183</td>
</tr>
<tr>
<td>N1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A24</td>
<td>1.152</td>
<td>-11.211</td>
</tr>
<tr>
<td>N28</td>
<td>7.784</td>
<td>-10.678</td>
</tr>
<tr>
<td>A15</td>
<td>0.886</td>
<td>-12.020</td>
</tr>
<tr>
<td>N37</td>
<td>0.546</td>
<td>-10.371</td>
</tr>
<tr>
<td>A11</td>
<td>4.294</td>
<td>-10.877</td>
</tr>
<tr>
<td>N43</td>
<td>1.434</td>
<td>-11.417</td>
</tr>
<tr>
<td>A29</td>
<td>2.029</td>
<td>-10.982</td>
</tr>
<tr>
<td>N23</td>
<td>9.485</td>
<td>-10.695</td>
</tr>
</tbody>
</table>
All items across both forms are now on the same scale. Because the IRT model estimates a person's ability (θ), all of the items may be explored as a single instrument. Even if all participants did not complete a particular item, the mathematical modeling of IRT and the subsequent equating procedure permits this activity because the model creates estimates of a person's Theta. The discussion of the items as a single instrument will be referred to as MCLCmr and has a 53 item pool.

An examination of MCLCmr revealed 27 items (NN49, N15, AA17, NN12, N11, A1, N21, N36, A23, A25, N42, N50, N51, N34, A20, AA14, A32, A22, N48, A26, A8, A30, A33, A27, N33, N4, N5) with unacceptable item discriminations, a parameters. Fifteen of these MCLCmr unacceptable item discriminations are negative (NN49, N15, AA17, NN12, N11, A1, N21, N36, A23, A25, N42, N50, N51, and N34). These items ranged from -0.021 to -2.639. The source of these negative discriminations is not known; however, these 15 items will not be moved to the next administration without revisions to the text or format.

Twenty-three MCLCmr items fell within the acceptable (≥0.3 and ≤2.5) discrimination range (NN17, N1, A31, N41, A16, A21, A28, N11, N37, N2, N44, A5, N6, N52, A15, N14, N47, N10, A24, A7, N43, NN46, A29). Three items had discrimination values greater than 2.5 (A11, N28, N23). The 23 items with acceptable discrimination values are evaluated relative to their difficulty parameters.

Six items in MCLCmr had positive b parameters; however, none of these six items had discrimination numbers within the acceptable range. The 47 items with negative values ranged from -0.833 to -37.578. These results suggest that the majority of
the items were exceptionally easy to correctly answer or guess. The 23 items within the acceptable discrimination range had multiple items falling within the same $b$ parameter range.

When selecting the items from the MCLCmr pool for the next administration, the item with the best discrimination level would inform the decisions. For example, two of the easiest items (NN17 and A15) both had $b$ values around -12.0; however, the discrimination numbers were 0.322 and 0.886, respectively. There would be no need to keep both items in the next administration because they are in essence measuring the same difficulty level. Item A15 would be the best choice to keep because as it would have a steeper slope as indicated in an ICC graph. It discriminates across a wider range of the latent trait.

Another issue related to the results of this study may be related to commonly held misconceptions. For example, 5% of the 573 participants correctly responded to the statement, *There is too little evidenced-based data to support the claim that parental involvement, such as reading to children, has a positive effect on children's reading acquisition.* Fourteen percent of the 503 participants correctly responded to the statement, *There is research-based evidence that storytelling is a strong predictor in later reading comprehension.* Considering the results of this study, alternative test formats might need to be explored. Constructing a multiple-choice format may improve the ability to have items with greater difficulties. The TRUE/FALSE format may be creating some odd residuals between people that are guessing and those that do not know the concepts. Additional administrations and additional item constructions are needed to enlarge the test pool to have items with a wider difficulty range that will assist in the discriminating people across a broader range of the latent trait, i.e., core knowledge of early literacy concepts.
CHAPTER 5

CONCLUSIONS

The primary objective of this work was the development of a research-based instrument (EL-Capstone) that may be used in future research investigations designed to measure the emergent literacy, core knowledge levels of adult practitioners working in an environment with expectations of supporting literacy development of young children. This instrument was informed by the processes used in the construction of concept inventories and recommended psychometric scale development methods.

In discussing the science of content specification, Webb points out that it is "based on expert judgments, writing effective test items, and balancing the many tradeoffs that have to be made" (2006, p. 155). The content domain was bound by the NELP experts and reduced content related reliability issues. Item construction from this first round of development has identified items that discriminated well on the latent ability; however, the analysis has also been successful in identifying items that will need to be discarded or revised. Baranowski (2006) cautions that during item editing and editorial review even minor changes to test items may impact the performance of the item.

IRT is designed to account for the covariance between items. As a non-linear model with structure, the approach to making decisions related to items that will move forward to additional administrations is not a simple statement that an item is good or bad. As a non-linear slope model, it is more of a decision related to asking for whom the item is good or bad. Each item has a maximum point of the slope or the 50/50 discrimination point. One of the next administration questions is to ask where on theta we want to be measuring.

The more reliable the instrument, the more precisely the latent construct, i.e., early literacy core knowledge, may be measured. Therefore, the standard error of measurement is lowered as the reliability is increased. Reliability is a matter of degree. While CTT depends on a corresponding single standard error of measurement, IRT precision is based on the latent trait being measured. Under the conditions of IRT, each
item contributes independently and allows the reliability to be described in terms of information independent of the total set of items or the sample participants as a group. It should be noted that this is not the case under CTT, where the total scores are influenced by other items and modifying any of the test items affects the total score.

Under IRT, the amount of information an item represents becomes a function of the amount of precision that item delivers. For example, if the item has high discrimination ability, that information is useful and is able to bring precision to distinguishing differences between high and low $\theta$. Basically, as the discrimination parameter increases, the information increases.

In EL-Capstone, the items with the difficulty parameter $(b)$ closest to $\theta$ and discrimination parameter $(a)$ relatively high contain the most useful information. As $\theta$ moves away from the difficulty parameter in either a positive or negative direction, the item has less information. This typically translates into where the slope of the ICC curve is the steepest.

This information may be helpful in future studies. In discussing item analysis, Livingston emphasizes "statistics alone cannot determine which items on a test are good and which are bad, but statistics can be used to identify items that are worth a particularly close look" (p. 423). This quality of information feature of IRT supports the items from this administration to be regrouped to form a new version of EL-Capstone for future administrations while still being able to anticipate the reliability of the instrument.

One of the next steps in the construction of EL-Capstone includes identifying the items that will move forward into the next test iteration. Being designed to measure early literacy knowledge levels across a wide range of ability, the next iterations of EL-Capstone may include some items that were identified as easy by the anonymous population in this study.

This initial phase of development has revealed gaps in the difficulty level of the item pool. The challenge for the next round of development will be to construct items that will have a higher difficulty level. Including a range of difficulty level items would still allow for discrimination at the lower end of the ability scale since there are adequate items identified as easy from the MCLCmr pool developed in this study. It should be noted, that a larger participant number in the online administration would have the
potential to impact both the a and the b parameters. Therefore, moving forward instrument items that allow participants across a wide early literacy knowledge levels to demonstrate some evidence of early literacy concepts minimizes floor and ceiling effects.

Investigations into the item placement location effects are another future research direction. For example, the use of multiple forms may be used with the same set of items. Form A would present the items starting with the items identified as easy by this first administration of EL-Capstone and progress to the more difficult items. Form B would reverse the order and begin with the more difficult items. Under the IRT assumptions of unidimensionality, the order effects should be minimum since the model is item focused. Early adaptive computer testing by Kingston and Dorans (1984) showed location effects in reading comprehension test items, analysis of explanations, and logical diagrams. While these finding may not be present in the True/False dichotomous format of EL-Capstone, additional investigations into item location effects may still be of interest.

Additional research needs include investigations into determining a particular "mastery" level of θ. In discussing content-related validity issues of latent variables measured through scales, Kane states "it is also necessary to evaluate the performance of the indicators as an observable attribute" (2006, p. 151). Future administrations of the instrument in well-defined ability sample groups may allow for item mapping of the response probability with observable matching. This mapping information may help determine what level of θ has the greatest impact on a child's long-term literacy gains. Creating anchor points in ability would be another meaningful research proposal. For example, defining what would be considered basic, proficient, and advanced core knowledge levels may support policy decisions related to the hiring and training of practitioners in environments with expectations of supporting early literacy.

In conclusion, budget cuts have affected public library (PL) services. At-risk children populations are vulnerable as operating-hour reductions force some libraries to lock doors in the evenings and weekends. Competition for funds has professionals seeking innovative ways to continue to provide important children services. An ability to demonstrate the quality, impact, and value associated with library services to long-term development of young children's conventional literacy skills is crucial when competing within limited funding resources. Research collaborations with public libraries are
expanding. Documenting EL core knowledge levels of adults that promote EL skills for young children is one innovative research-based approach to demonstrate quality of services.

The proposed design of this study has been completed. The items, as identified through the IRT analysis, are ready to move to the next round of revisions, administration, and analysis that is required to make EL-Capstone a valuable tool for use in environments with expectations of nurturing early literacy development in young children.

Some concept inventories have evolved through additional testing of the instrument to be used to measure learning outcomes. While EL-Capstone may be expanded in the future for use in measuring learning outcomes, this instrument would still need to be tested and calibrated before being expanded since it has not been validated in this area. Perhaps this research, combined with other similar research agendas, will expand the initial application to include the assessment of learning outcomes, the evaluation of teaching methods, and the informing of curriculum design.
APPENDIX A

RECRUITMENT LETTERS

Librarian Participant Mailed/E-mailed Recruitment Script Sample

Participant Address Here
Dear [participant name],

Greetings! I am a doctoral candidate with the College of Communication & Information Studies at the Florida State University conducting a research project. Libraries have historically been making vital connections between children, books, and literacy. Funding initiatives demonstrate that libraries are developing partnerships to expanded children services to include emergent literacy components. The goal of this research is to design a valid and reliable, community-sensitive tool to assess emergent literacy understanding in adults that may be called upon as leaders to promote emergent literacy development of young children.

As a professional librarian, your perspective will be helpful to this study. I would like to solicit your support by inviting you to participate in a 45-60 minute interview as you complete a prototype of this emergent literacy instrument.

I know your time is valuable, but by taking time out of your busy schedule to participate in this study, your contribution will benefit libraries and the direction of emergent literacy leadership preparedness of librarians that work with young children.

If you would like to volunteer to participate in this study and/or you would like me to provide more background on the study, please email me at jcapps@fsu.edu.

Your participation in this research project is COMPLETELY VOLUNTARY and you may quit at any time without any penalty to you. All of your responses will be KEPT CONFIDENTIAL and any data collected from this study will be reported in aggregate form.

I look forward to our future collaboration on this project.

Thank you.

Sincerely,
Janet Capps

College of Communication & Information

107
Greetings! I am a doctoral candidate with the College of Communication & Information Studies at the Florida State University conducting a research project. Libraries have historically made vital connections between children, books, and literacy. Funding initiatives demonstrate that libraries are developing partnerships to expanded children services to include emergent literacy components. The goal of this research is to design a valid and reliable, community-sensitive tool to assess emergent literacy understanding in adults that may be called upon as leaders to promote emergent literacy development of young children.

As an adult over 18 years of age, your participation will be helpful to this study. I would like to solicit your support by inviting you to participate in a pilot test of an emergent literacy instrument. Your volunteer task will be to complete an online version of the emergent literacy instrument (approximately 30 minutes). No experience in emergent literacy concepts is required. Respondent participation is vital to the item selection process of this instrument. Your scores will be kept confidential, and will be used only to evaluate this emergent literacy concept inventory known as EL-Capstone.

I know your time is valuable, but by taking time out of your busy schedule to participate in this study, your contribution will benefit libraries and the direction of emergent literacy leadership preparedness of librarians and adults that work with young children.

If you would like to volunteer to participate in this research, and/or you would like me to provide more background on the study, please email me at jcapps@fsu.edu.

Your participation in this research project is COMPLETELY VOLUNTARY and you may quit at any time without any penalty to you. All of your responses will be KEPT CONFIDENTIAL and any data collected from this study will be reported in aggregate form.

I look forward to our future collaboration on this project.

Thank you.

Sincerely,
Janet Capps

School of Communication & Information

[Contact information redacted]
APPENDIX B

CONSENT FORMS

EL-Capstone: Emergent Literacy Concept Inventory
Informed Consent Form (Librarians)

You are invited to participate in a research project designed to develop an instrument to measure emergent literacy capstone levels of adults. The duration of this project will be April 2010 through December 2010. This study is being conducted by Janet L. Capps, College of Communication & Information, The Florida State University.

You were selected as a possible participant because you were identified as a professional librarian. I ask that you read this form and ask any questions you may have before agreeing to be in the study.

Libraries have historically made vital connections between children, books, and literacy. Funding initiatives demonstrate libraries are developing partnerships to expanded children services to include emergent literacy concepts. How knowledgeable are librarians in core emergent literacy concepts? Having a valid and reliable, community-sensitive tool to measure the capstone levels at which librarians are prepared to fulfill a leadership role in the community's effort to impact emergent literacy development is paramount to future design and implementation of librarian training intervention programs and library information science education courses.

Procedures consistent with sound research design are in place. We know of no risks associated with your participation in this study. However, in the unlikely event of unforeseen discomfort associated or involved in this study, your participation may stop immediately. Participation in this study is voluntary. There is no compensation for your participation in this study; however, expanded understanding of emergent literacy concepts may be a direct benefit from participation. Your decision whether or not to participate will not affect your current or future relations with the Florida State University. You are free to not answer any question or withdraw at any time without affecting those relationships.

If you choose to participate in the study, you will be interviewed as you complete the EL-Capstone, an instrument being designed to measure emergent literacy capstone levels of adults. Data gathered from this interview will be used to evaluate and inform the EL-Capstone’s test item design components. Your scores are kept confidential.

All information obtained as part of this project will be kept private and confidential to the extent permitted by law. Your name will not be included in any of the reports describing this project or disseminating its findings. Data collected as part of this project will be kept on a secure computer (on password protected files) until December 2014 under the direction of Dr. Kathleen Burnet.

If you have any questions now or at any time during the duration of the study, you may contact Janet L. Capps, [contact information]. An alternative contact is the research advisor, Dr. Kathleen Burnet [contact information].

If you have any questions or concerns regarding this study and would like to talk to someone other than the researcher, you are encouraged to contact the FSU IRB at 100 Levy Street, Research Building E, Suite 276, Tallahassee, FL 32306-3743, or 850-644-6533, or by email at humanobjects@magseis.fsu.edu.

If you agree to participate in this research project, please sign and print your name. Your signature indicates you are over the age of 18, have read the information provided above, and that you freely and voluntarily, without element of coercion, agree to participate in this research project. A copy of this consent form has been offered to you; additional copies are available upon request.

Printed name

Signature

Today’s date

Work Address

EL-Capstone: Emergent Literacy Concept Inventory
Informed Consent Form (Adult participant)

You are invited to participate in a research project designed to develop an instrument to measure emergent literacy capstone levels of adults. The duration of this project will be April 2010 through December 2010. This study is being conducted by Janet L. Capps, College of Communication & Information, The Florida State University. I ask that you read this form and ask any questions you may have before agreeing to participate in the study.

Libraries have historically made vital connections between children, books, and literacy. Funding initiatives demonstrate libraries are developing partnerships to expanded children services to include emergent literacy concepts. How knowledgeable are librarians in core emergent literacy concepts? Having a valid and reliable, community-sensitive tool to measure the capstone levels at which librarians are prepared to fulfill a leadership role in the community’s effort to impact emergent literacy development is paramount to future design and implementation of library training intervention programs and library information science education courses.

Procedures consistent with sound research design are in place. We know of no risks associated with your participation in this study. However, in the unlikely event of unforeseen discomfort associated or involved in this study, your participation may stop immediately. Participation in this study is voluntary. There is no compensation for your participation in this study; however, expanded understanding of emergent literacy concepts may be a direct benefit from participation. Your decision whether or not to participate will not affect your current or future relations with the Florida State University. You are free to not answer any question or withdraw at any time without affecting those relationships.

If you choose to participate in the study, you will be asked to complete an online version of EL-Capstone, an instrument designed to measure emergent literacy capstone levels of adults. Data gathered from these tests will be used to evaluate and inform the EL-Capstone’s test items design components. Respondent participation is vital to the item selection process of this instrument. Your scores are kept confidential.

All information obtained as part of this project will be kept private and confidential to the extent permitted by law. Your name will not be included in any of the reports describing this project or disseminating its findings. Data collected as part of this project will be kept on a secure computer (in password protected files) until December 2014 under the direction of Dr. Kathleen Burner.

If you have any questions now or at anytime during the duration of the study, you may contact Janet L. Capps, [contact information]. An alternative contact is the research advisor, Dr. Kathleen Burner [contact information].

If you have any questions or concerns regarding this study and would like to talk to someone other than the researcher, you are encouraged to contact the FSU IRB at 301 Levy Street, Research Building B, Suite 276, Tallahassee, FL 32306-1742, or 850-644-8633, or by email at humansubjects@magnet.fsu.edu.

Clicking ‘Yes’ indicates you are over the age of 18, have read the information provided above, and that you freely and voluntarily, without element of coercion, agree to participate in this research project. A copy of this consent form may be printed for your records; alternative file formats are available upon request.

Question 1. Please indicate your agreement with the consent form.

☐ Yes
☐ No

FSU Human Subject: Committee Approved 3/14/10 Void after 11/8/10. HSC# 2010.4524
APPENDIX C

INTERVIEW DATA

The results of the "N" unique item supplemental questions collected during the interview sessions (0=lowest ranking and 6= highest ranking).

<table>
<thead>
<tr>
<th>ID#</th>
<th>Early Literacy Concept?</th>
<th>Adult Ability to Answer</th>
<th>Important Adult Concept</th>
</tr>
</thead>
<tbody>
<tr>
<td>N1</td>
<td>6</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>N2</td>
<td>4</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>N3</td>
<td>6</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>N4</td>
<td>6</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>N5</td>
<td>6</td>
<td>3</td>
<td>6</td>
</tr>
<tr>
<td>N6</td>
<td>5</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>N7</td>
<td>6</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>N8</td>
<td>6</td>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td>N9</td>
<td>6</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>N10</td>
<td>6</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>N11</td>
<td>6</td>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td>N12</td>
<td>4</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>N13</td>
<td>6</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>N14</td>
<td>6</td>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td>N15</td>
<td>6</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>N16</td>
<td>6</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>N17</td>
<td>6</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>N18</td>
<td>5</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>N19</td>
<td>3</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>N20</td>
<td>5</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>N21</td>
<td>1</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>N22</td>
<td>6</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>N23</td>
<td>6</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>N24</td>
<td>5</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>N25</td>
<td>5</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>N26</td>
<td>6</td>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td>N27</td>
<td>6</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>N28</td>
<td>6</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>N29</td>
<td>6</td>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td>N30</td>
<td>6</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>N31</td>
<td>5</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>N32</td>
<td>6</td>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td>N33</td>
<td>3</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>ID#</td>
<td>Early Literacy Concept?</td>
<td>Adult Ability to Answer</td>
<td>Important Adult Concept</td>
</tr>
<tr>
<td>-------</td>
<td>-------------------------</td>
<td>-------------------------</td>
<td>------------------------</td>
</tr>
<tr>
<td>N34</td>
<td>6</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>N35</td>
<td>6</td>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td>N36</td>
<td>5</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>N37</td>
<td>5</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>N38</td>
<td>6</td>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td>N39</td>
<td>4</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>N40</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>N41</td>
<td>6</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>N42</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>N43</td>
<td>6</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>N44</td>
<td>5</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>N45</td>
<td>6</td>
<td>4</td>
<td>6</td>
</tr>
</tbody>
</table>

The results of the "A" unique item supplemental questions collected during the interview sessions (0=lowest ranking and 6= highest ranking).

<table>
<thead>
<tr>
<th>ID#</th>
<th>Early Literacy Concept?</th>
<th>Adult Ability to Answer</th>
<th>Important Adult Concept</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>A2</td>
<td>4</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>A3</td>
<td>6</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>A4</td>
<td>6</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>A5</td>
<td>6</td>
<td>2</td>
<td>6</td>
</tr>
<tr>
<td>A6</td>
<td>6</td>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td>A7</td>
<td>6</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>A8</td>
<td>6</td>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td>A9</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>A10</td>
<td>6</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>A11</td>
<td>6</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>A12</td>
<td>6</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>A13</td>
<td>6</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>A14</td>
<td>6</td>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td>A15</td>
<td>6</td>
<td>3</td>
<td>6</td>
</tr>
<tr>
<td>A16</td>
<td>6</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>A17</td>
<td>6</td>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td>A18</td>
<td>6</td>
<td>3</td>
<td>6</td>
</tr>
<tr>
<td>A19</td>
<td>6</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>A20</td>
<td>6</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>A21</td>
<td>5</td>
<td>3</td>
<td>5</td>
</tr>
</tbody>
</table>
The results of the "A" common item supplemental questions collected during ten interview sessions (0=lowest ranking and 6= highest ranking).

<table>
<thead>
<tr>
<th>ID#</th>
<th>Early Literacy Concept?</th>
<th>Adult Ability to Answer</th>
<th>Important Adult Concept</th>
</tr>
</thead>
<tbody>
<tr>
<td>A22</td>
<td>5</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>A23</td>
<td>6</td>
<td>3</td>
<td>5</td>
</tr>
</tbody>
</table>
APPENDIX D
RECRUITMENT POSTING EXAMPLES

Capstone Study: We invite you to join us

Foundation for Early Learning is delighted to feature a guest blog post from Janet Capps, a member of the Early Learning Public Library Partnership. If you are interested in being featured in the Foundation’s blog, please contact Joel Balestra, Communications Manager at jde@earlylearning.org.

Starting up any new learning opportunities or making advancements in the field of early learning for children is a collaborative effort that includes academic research.

If you are an adult over 18 years of age, you are invited to participate in a research project that has the potential to impact the development of a new tool (research-based, community-sensitive) that is being constructed under the guidance of the Florida State University.

To participate or to learn more, please visit: http://www.elcapstone.com/

Early Literacy Survey

September 9th, 2010 | Posted by Dr. Eddy Damon Cohen in early literacy | +1 Share/Bookmark

A colleague who is a doctoral candidate at Florida State University is working on an adult early literacy assessment tool, called EL-Capstone that will be used to assess the understanding of early literacy concepts of adults who may be called upon as leaders to promote early literacy. She is looking for 1000 anonymous volunteers that spread across a wide range of early literacy concept awareness levels to take the sample survey. Data collection for the this early version of the EL-Capstone instrument has already launched. Below is the blurb describing the project, since instrument development cannot happen in isolation and needs the support of many people, I encourage you to participate and send the blurb to your colleagues.

Anonymous volunteers needed to support early literacy academic research in an online activity.

Making advancements in the field of early learning for children is a collaborative effort that includes academic research. If you are an adult over 18 years of age, you are invited to participate in a research project and to impact the development of a new tool that is being constructed under the guidance of the Florida State University. When finished, this tool will be used to assess the understanding of early literacy concepts of adults who may be called upon as leaders to promote early literacy development of young children. No experience or background in early literacy concepts is required. To participate, please visit: http://www.elcapstone.com
Note. Written permission has been obtained for all screen captures in this appendix.
APPENDIX E

PARAMETERS (A, RAW B, AND B)

The 2-parameters, i.e., $a$ and $b$, are presented for both forms BC and LC. The raw-b value indicates $b$ prior to calculating the IRT parameterization in a 2-PL metric where the logit (natural log of the odds ratio) is $1.7 \times$ discrimination $\times$ (theta - difficulty). The True/False participant response percentages are shown for each item in forms BC and LC.

Form Big C (BC)

<table>
<thead>
<tr>
<th>Item</th>
<th>$a$</th>
<th>raw-b</th>
<th>$b$</th>
<th>Response % True</th>
<th>Response % False</th>
</tr>
</thead>
<tbody>
<tr>
<td>A31</td>
<td>0.567</td>
<td>-1.778</td>
<td>-3.136</td>
<td>84</td>
<td>16</td>
</tr>
<tr>
<td>N47</td>
<td>1.598</td>
<td>-2.403</td>
<td>-1.504</td>
<td>84</td>
<td>16</td>
</tr>
<tr>
<td>A20</td>
<td>0.059</td>
<td>2.906</td>
<td>49.254</td>
<td>5</td>
<td>95</td>
</tr>
<tr>
<td>N5</td>
<td>0.498</td>
<td>-2.301</td>
<td>-4.620</td>
<td>90</td>
<td>10</td>
</tr>
<tr>
<td>N4</td>
<td>0.472</td>
<td>-2.192</td>
<td>-4.644</td>
<td>89</td>
<td>11</td>
</tr>
<tr>
<td>A30</td>
<td>0.326</td>
<td>-2.532</td>
<td>-7.767</td>
<td>91</td>
<td>9</td>
</tr>
<tr>
<td>N11</td>
<td>0.888</td>
<td>-2.469</td>
<td>-2.780</td>
<td>90</td>
<td>10</td>
</tr>
<tr>
<td>A27</td>
<td>0.418</td>
<td>-2.437</td>
<td>-5.830</td>
<td>91</td>
<td>9</td>
</tr>
<tr>
<td>N41</td>
<td>0.704</td>
<td>-3.166</td>
<td>-4.497</td>
<td>95</td>
<td>5</td>
</tr>
<tr>
<td>N1</td>
<td>0.548</td>
<td>-2.863</td>
<td>-5.224</td>
<td>95</td>
<td>5</td>
</tr>
<tr>
<td>A5</td>
<td>1.107</td>
<td>-0.989</td>
<td>-0.893</td>
<td>70</td>
<td>30</td>
</tr>
<tr>
<td>N14</td>
<td>1.549</td>
<td>-2.735</td>
<td>-1.766</td>
<td>88</td>
<td>12</td>
</tr>
<tr>
<td>A26</td>
<td>0.275</td>
<td>-3.086</td>
<td>-11.222</td>
<td>95</td>
<td>5</td>
</tr>
<tr>
<td>N33</td>
<td>0.427</td>
<td>-1.152</td>
<td>-2.698</td>
<td>75</td>
<td>25</td>
</tr>
<tr>
<td>A28</td>
<td>0.78</td>
<td>-1.13</td>
<td>-1.449</td>
<td>74</td>
<td>26</td>
</tr>
<tr>
<td>N34</td>
<td>0.036</td>
<td>-0.846</td>
<td>-23.500</td>
<td>70</td>
<td>30</td>
</tr>
<tr>
<td>A16</td>
<td>0.726</td>
<td>-3.588</td>
<td>-4.942</td>
<td>97</td>
<td>3</td>
</tr>
<tr>
<td>N44</td>
<td>1.013</td>
<td>-1.927</td>
<td>-1.902</td>
<td>84</td>
<td>16</td>
</tr>
<tr>
<td>N52</td>
<td>1.289</td>
<td>-1.527</td>
<td>-1.185</td>
<td>76</td>
<td>24</td>
</tr>
<tr>
<td>A33</td>
<td>0.376</td>
<td>-2.836</td>
<td>-7.543</td>
<td>94</td>
<td>6</td>
</tr>
<tr>
<td>N50</td>
<td>0.135</td>
<td>-1.149</td>
<td>-8.511</td>
<td>77</td>
<td>23</td>
</tr>
<tr>
<td>A1</td>
<td>0.773</td>
<td>-0.823</td>
<td>-1.065</td>
<td>33</td>
<td>67</td>
</tr>
<tr>
<td>N48</td>
<td>0.261</td>
<td>-1.187</td>
<td>-4.548</td>
<td>24</td>
<td>76</td>
</tr>
<tr>
<td>A25</td>
<td>0.201</td>
<td>0.636</td>
<td>3.164</td>
<td>67</td>
<td>33</td>
</tr>
<tr>
<td>N16</td>
<td>0.811</td>
<td>0.676</td>
<td>0.834</td>
<td>65</td>
<td>35</td>
</tr>
<tr>
<td>N21</td>
<td>0.433</td>
<td>0.894</td>
<td>2.065</td>
<td>71</td>
<td>29</td>
</tr>
<tr>
<td>N42</td>
<td>0.19</td>
<td>-1.449</td>
<td>-7.626</td>
<td>19</td>
<td>81</td>
</tr>
<tr>
<td>A8</td>
<td>0.32</td>
<td>-2.498</td>
<td>-7.806</td>
<td>8</td>
<td>92</td>
</tr>
<tr>
<td>A23</td>
<td>0.407</td>
<td>-1.549</td>
<td>-3.806</td>
<td>19</td>
<td>81</td>
</tr>
<tr>
<td>N51</td>
<td>0.048</td>
<td>1.816</td>
<td>37.833</td>
<td>86</td>
<td>14</td>
</tr>
</tbody>
</table>
Form Little c (LC)

<table>
<thead>
<tr>
<th>Item</th>
<th>a</th>
<th>raw-b</th>
<th>b</th>
<th>True</th>
<th>False</th>
</tr>
</thead>
<tbody>
<tr>
<td>N2</td>
<td>0.309</td>
<td>-1.503</td>
<td>-4.864</td>
<td>81</td>
<td>19</td>
</tr>
<tr>
<td>A32</td>
<td>0.069</td>
<td>-0.532</td>
<td>-7.710</td>
<td>62</td>
<td>38</td>
</tr>
<tr>
<td>N6</td>
<td>0.375</td>
<td>-1.829</td>
<td>-4.877</td>
<td>86</td>
<td>14</td>
</tr>
<tr>
<td>A21</td>
<td>0.252</td>
<td>-0.915</td>
<td>-3.631</td>
<td>69</td>
<td>31</td>
</tr>
<tr>
<td>N10</td>
<td>0.581</td>
<td>-1.182</td>
<td>-2.034</td>
<td>74</td>
<td>26</td>
</tr>
<tr>
<td>A7</td>
<td>0.721</td>
<td>-3.118</td>
<td>-4.325</td>
<td>95</td>
<td>5</td>
</tr>
<tr>
<td>N15</td>
<td>-1.039</td>
<td>1.672</td>
<td>-1.609</td>
<td>20</td>
<td>80</td>
</tr>
<tr>
<td>A22</td>
<td>-0.069</td>
<td>0.647</td>
<td>-9.377</td>
<td>33</td>
<td>67</td>
</tr>
<tr>
<td>N36</td>
<td>-0.14</td>
<td>-0.346</td>
<td>2.471</td>
<td>57</td>
<td>43</td>
</tr>
<tr>
<td>N1</td>
<td>0.652</td>
<td>-2.525</td>
<td>-3.873</td>
<td>91</td>
<td>9</td>
</tr>
<tr>
<td>A5</td>
<td>0.423</td>
<td>-1.033</td>
<td>-2.442</td>
<td>73</td>
<td>27</td>
</tr>
<tr>
<td>A24</td>
<td>0.639</td>
<td>-2.388</td>
<td>-3.737</td>
<td>90</td>
<td>10</td>
</tr>
<tr>
<td>N28</td>
<td>4.319</td>
<td>-9.079</td>
<td>-2.102</td>
<td>98</td>
<td>2</td>
</tr>
<tr>
<td>A15</td>
<td>0.491</td>
<td>-3.054</td>
<td>-6.220</td>
<td>95</td>
<td>5</td>
</tr>
<tr>
<td>N37</td>
<td>-0.303</td>
<td>0.352</td>
<td>-1.162</td>
<td>41</td>
<td>59</td>
</tr>
<tr>
<td>A11</td>
<td>2.382</td>
<td>-6.458</td>
<td>-2.711</td>
<td>99</td>
<td>1</td>
</tr>
<tr>
<td>N43</td>
<td>0.796</td>
<td>-3.476</td>
<td>-4.367</td>
<td>95</td>
<td>5</td>
</tr>
<tr>
<td>A29</td>
<td>1.126</td>
<td>-3.418</td>
<td>-3.036</td>
<td>95</td>
<td>5</td>
</tr>
<tr>
<td>N23</td>
<td>5.262</td>
<td>-11.337</td>
<td>-2.155</td>
<td>98</td>
<td>2</td>
</tr>
<tr>
<td>AA17</td>
<td>-0.56</td>
<td>1.661</td>
<td>-2.966</td>
<td>84</td>
<td>16</td>
</tr>
<tr>
<td>NN46</td>
<td>0.848</td>
<td>-3.751</td>
<td>-4.423</td>
<td>4</td>
<td>96</td>
</tr>
<tr>
<td>NN12</td>
<td>-0.539</td>
<td>0.96</td>
<td>-1.781</td>
<td>73</td>
<td>27</td>
</tr>
<tr>
<td>AA14</td>
<td>0.176</td>
<td>-0.973</td>
<td>-5.528</td>
<td>28</td>
<td>72</td>
</tr>
<tr>
<td>NN17</td>
<td>0.66</td>
<td>-2.337</td>
<td>-3.541</td>
<td>10</td>
<td>90</td>
</tr>
<tr>
<td>N16</td>
<td>-0.555</td>
<td>0.87</td>
<td>-1.568</td>
<td>70</td>
<td>30</td>
</tr>
<tr>
<td>N21</td>
<td>0.022</td>
<td>1.248</td>
<td>56.727</td>
<td>77</td>
<td>23</td>
</tr>
<tr>
<td>N42</td>
<td>0.179</td>
<td>-1.318</td>
<td>-7.363</td>
<td>21</td>
<td>79</td>
</tr>
<tr>
<td>A8</td>
<td>0.052</td>
<td>-2.231</td>
<td>-42.904</td>
<td>10</td>
<td>90</td>
</tr>
<tr>
<td>NN49</td>
<td>-1.464</td>
<td>4.162</td>
<td>-2.843</td>
<td>97</td>
<td>3</td>
</tr>
</tbody>
</table>
# APPENDIX F

## EQUATING COEFFICIENTS

### Form Big C (BC)

<table>
<thead>
<tr>
<th>Item</th>
<th>a</th>
<th>raw-b</th>
<th>b</th>
</tr>
</thead>
<tbody>
<tr>
<td>N1</td>
<td>0.548</td>
<td>-2.863</td>
<td>-5.224</td>
</tr>
<tr>
<td>A5</td>
<td>1.107</td>
<td>-0.989</td>
<td>-0.893</td>
</tr>
<tr>
<td>N16</td>
<td>0.811</td>
<td>0.676</td>
<td>0.834</td>
</tr>
<tr>
<td>N21</td>
<td>0.433</td>
<td>0.894</td>
<td>2.065</td>
</tr>
<tr>
<td>N42</td>
<td>0.19</td>
<td>-1.449</td>
<td>-7.626</td>
</tr>
<tr>
<td>A8</td>
<td>0.32</td>
<td>-2.498</td>
<td>-7.806</td>
</tr>
</tbody>
</table>

- mean a 0.568
- mean b -3.109
- std b 4.340

### Form Little c (LC)

<table>
<thead>
<tr>
<th>Item</th>
<th>a</th>
<th>raw-b</th>
<th>b</th>
</tr>
</thead>
<tbody>
<tr>
<td>N1</td>
<td>0.652</td>
<td>-2.525</td>
<td>-3.873</td>
</tr>
<tr>
<td>A5</td>
<td>0.423</td>
<td>-1.033</td>
<td>-2.442</td>
</tr>
<tr>
<td>N16</td>
<td>-0.555</td>
<td>0.87</td>
<td>-1.568</td>
</tr>
<tr>
<td>N21</td>
<td>0.022</td>
<td>1.248</td>
<td>56.727</td>
</tr>
<tr>
<td>N42</td>
<td>0.179</td>
<td>-1.318</td>
<td>-7.363</td>
</tr>
<tr>
<td>A8</td>
<td>0.052</td>
<td>-2.231</td>
<td>-42.904</td>
</tr>
</tbody>
</table>

- mean a 0.129
- mean b -0.237
- std b 32.050
## APPENDIX G

### NEW PARAMETERS FORM LITTLE C

<table>
<thead>
<tr>
<th>Item</th>
<th>a</th>
<th>b</th>
</tr>
</thead>
<tbody>
<tr>
<td>N2</td>
<td>2.282</td>
<td>-0.896</td>
</tr>
<tr>
<td>A32</td>
<td>0.510</td>
<td>-1.281</td>
</tr>
<tr>
<td>N6</td>
<td>2.769</td>
<td>-0.897</td>
</tr>
<tr>
<td>A21</td>
<td>1.861</td>
<td>-0.729</td>
</tr>
<tr>
<td>N10</td>
<td>4.291</td>
<td>-0.512</td>
</tr>
<tr>
<td>A7</td>
<td>5.325</td>
<td>-0.823</td>
</tr>
<tr>
<td>N15</td>
<td>-7.673</td>
<td>-0.455</td>
</tr>
<tr>
<td>A22</td>
<td>-0.510</td>
<td>-1.507</td>
</tr>
<tr>
<td>N36</td>
<td>-1.034</td>
<td>0.098</td>
</tr>
<tr>
<td>N1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A24</td>
<td>4.719</td>
<td>-0.743</td>
</tr>
<tr>
<td>N28</td>
<td>31.897</td>
<td>-0.522</td>
</tr>
<tr>
<td>A15</td>
<td>3.626</td>
<td>-1.079</td>
</tr>
<tr>
<td>N37</td>
<td>-2.238</td>
<td>-0.394</td>
</tr>
<tr>
<td>A11</td>
<td>17.592</td>
<td>-0.604</td>
</tr>
<tr>
<td>N43</td>
<td>5.879</td>
<td>-0.828</td>
</tr>
<tr>
<td>A29</td>
<td>8.316</td>
<td>-0.648</td>
</tr>
<tr>
<td>N23</td>
<td>38.861</td>
<td>-0.529</td>
</tr>
<tr>
<td>AA17</td>
<td>-4.136</td>
<td>-0.639</td>
</tr>
<tr>
<td>NN46</td>
<td>6.263</td>
<td>-0.836</td>
</tr>
<tr>
<td>NN12</td>
<td>-3.981</td>
<td>-0.478</td>
</tr>
<tr>
<td>AA14</td>
<td>1.300</td>
<td>-0.986</td>
</tr>
<tr>
<td>NN17</td>
<td>4.874</td>
<td>-0.716</td>
</tr>
<tr>
<td>N16</td>
<td></td>
<td></td>
</tr>
<tr>
<td>N21</td>
<td></td>
<td></td>
</tr>
<tr>
<td>N42</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NN49</td>
<td>-10.812</td>
<td>-0.622</td>
</tr>
</tbody>
</table>
APPENDIX H

ONLINE ADMINISTRATION DATA

The ICC for the N-items using the adjusted IRT b parameters and the transformation coefficients for converting the LC items into the scale of the BC. These curves are pre-reverse coding that accounted for items with negated wording.

Item N1-Cc

<table>
<thead>
<tr>
<th>Item</th>
<th>a</th>
<th>b</th>
</tr>
</thead>
<tbody>
<tr>
<td>N1</td>
<td>0.548</td>
<td>-5.22445</td>
</tr>
</tbody>
</table>

Item N2-c

<table>
<thead>
<tr>
<th>Item</th>
<th>a</th>
<th>b</th>
</tr>
</thead>
<tbody>
<tr>
<td>N2</td>
<td>2.28206</td>
<td>-0.89563</td>
</tr>
</tbody>
</table>
Item N4-C

Item N5-C

Item N6-c
Item N17-c

Item N21-Cc

Item N23-c

<table>
<thead>
<tr>
<th>Item</th>
<th>a</th>
<th>b</th>
</tr>
</thead>
<tbody>
<tr>
<td>N17</td>
<td>4.87</td>
<td>-0.72</td>
</tr>
<tr>
<td>N21</td>
<td>0.43</td>
<td>2.06</td>
</tr>
<tr>
<td>N23</td>
<td>38.86</td>
<td>-0.53</td>
</tr>
</tbody>
</table>
Item N28-c

Item N33-C

Item N34-C

Item   a     b
N28    31.8971 -0.52164

Item   a     b
N33    0.427 -2.69789

Item   a     b
N34    0.036 -23.5
Item N36-c

```
Item       a       b
N36    -1.03394  0.09763
```

Item N37-c

```
Item       a       b
N37    -2.23775  -0.39431
```

Item N41-C

```
Item       a       b
N41     0.704    -4.49716
```
Item N46-c

<table>
<thead>
<tr>
<th>Item</th>
<th>a</th>
<th>b</th>
</tr>
</thead>
<tbody>
<tr>
<td>N46</td>
<td>6.26274</td>
<td>-0.83595</td>
</tr>
</tbody>
</table>

Item N47-C

<table>
<thead>
<tr>
<th>Item</th>
<th>a</th>
<th>b</th>
</tr>
</thead>
<tbody>
<tr>
<td>N47</td>
<td>1.598</td>
<td>-1.50375</td>
</tr>
</tbody>
</table>

Item N48-C

<table>
<thead>
<tr>
<th>Item</th>
<th>a</th>
<th>b</th>
</tr>
</thead>
<tbody>
<tr>
<td>N48</td>
<td>0.261</td>
<td>-4.54789</td>
</tr>
</tbody>
</table>
Item N49-c

Item N50-C

Item N51-C

<table>
<thead>
<tr>
<th>Item</th>
<th>a</th>
<th>b</th>
</tr>
</thead>
<tbody>
<tr>
<td>N49</td>
<td>-10.8121</td>
<td>-0.62195</td>
</tr>
<tr>
<td>N50</td>
<td>0.135</td>
<td>-8.51111</td>
</tr>
<tr>
<td>N51</td>
<td>0.048</td>
<td>37.8333</td>
</tr>
</tbody>
</table>
The ICC for the A-items using the adjusted IRT b parameters and the transformation coefficients for converting the LC items into the scale of the BC.

**Item N52-C**

<table>
<thead>
<tr>
<th>Item</th>
<th>a</th>
<th>b</th>
</tr>
</thead>
<tbody>
<tr>
<td>N52</td>
<td>1.289</td>
<td>-1.18464</td>
</tr>
</tbody>
</table>

**Item A1-C**

<table>
<thead>
<tr>
<th>Item</th>
<th>a</th>
<th>b</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1</td>
<td>0.773</td>
<td>-1.06468</td>
</tr>
</tbody>
</table>
Item A5-Cc

Item A7-c

Item A8-Cc

Item  | a    | b      
-----|------|--------
A5   | 1.107| -0.89341 |
A7   | 5.32481 | -0.82257 |
A8   | 0.32 | -7.80625 |
Item A11-c

Probabilty

Theta

Item A14-c

Probabilty

Theta

Item A15-c

Probabilty

Theta

Item a b
A11 17.5918 -0.60411
A14 1.29981 -0.98558
A15 3.62619 -1.07922
Item A21-c

Item A22-c

Item A23-C

<table>
<thead>
<tr>
<th>Item</th>
<th>a</th>
<th>b</th>
</tr>
</thead>
<tbody>
<tr>
<td>A21</td>
<td>1.8611</td>
<td>-0.72866</td>
</tr>
<tr>
<td>A22</td>
<td>-0.50959</td>
<td>-1.50667</td>
</tr>
<tr>
<td>A23</td>
<td>0.407</td>
<td>-3.8059</td>
</tr>
</tbody>
</table>
Item A27-C

Item A28-C

Item A29-c

<table>
<thead>
<tr>
<th>Item</th>
<th>a</th>
<th>b</th>
</tr>
</thead>
<tbody>
<tr>
<td>A27</td>
<td>0.418</td>
<td>-5.83014</td>
</tr>
<tr>
<td>A28</td>
<td>0.78</td>
<td>-1.44872</td>
</tr>
<tr>
<td>A29</td>
<td>8.31585</td>
<td>-0.64803</td>
</tr>
</tbody>
</table>
Item A33-C

<table>
<thead>
<tr>
<th>Item</th>
<th>a</th>
<th>b</th>
</tr>
</thead>
<tbody>
<tr>
<td>A33</td>
<td>0.376</td>
<td>-7.54255</td>
</tr>
</tbody>
</table>

Theta

Probability

-3 -2.5 -2 -1.5 -1 -0.5 0 0.5 1 1.5 2 2.5 3
APPENDIX I

REVERSED CODED ICC

BCmr ICCs

LCmr ICCs
REFERENCES


150


BIOGRAPHICAL SKETCH

Janet Capps earned a Bachelor of Science degree (1997) and a Master of Science degree (2002) from the Florida State University College of Communication & Information. This dissertation submitted to the College of Communication & Information is in partial fulfillment of the requirements for the degree of Doctor of Philosophy to be awarded in April 2011.

Prior to returning to the Florida State University to start the doctoral program, she was a faculty member with the Media Studies Department at Radford University (2002-2004). She has taught beginning and advanced courses in electronic multimedia design & production, project management, and information organization. Janet has served as a research assistant for a National Science Foundation grant, a Department of Education Comprehensive School Reform grant, and an Institute for Museum and Library Services National Leadership grant.

Her research interests are centered on technology innovations in information delivery/retrieval and learning environments and include the areas of literacy and literacy behaviors (i.e., information, technology, and early childhood), information/knowledge inventories, information processing and eye movement control, and Library and Information Science education. Janet has served as President of Beta Phi Mu, Gamma Chapter, an editorial assistant/project manager for the Journal of Education for Library and Information Science, and as a College Representative for the Congress of Graduate Students.