

ARGUING FOR VALIDITY WHEN USING A SMALL SCALE STUDY TO DETERMINE
THE CONTENT OF A MONTESSORI TEACHER CREDENTIALING TEST

Diane M. Talley

A dissertation submitted to the faculty at the University of North Carolina at Chapel in partial fulfillment of the requirements for the degree of Doctor of Philosophy in School of Education.

Chapel Hill
2016

Approved by:

Rita O'Sullivan

Chad Buckendahl

Sherick Hughes

Stephen Johnson

William Ware

©2016
Diane M. Talley
ALL RIGHTS RESERVED

ABSTRACT

Measuring competency of a Montessori teacher for the purpose of credentialing requires a deep understanding of the role of these teachers in the classroom and school environment. In the credentialing industry, large-scale studies are conducted in order to determine the essential tasks performed on the job and the critical knowledge, skills, and abilities required for competent performance. Literature suggests a strong reliance on a single method for determining content of a credentialing test (Raymond, 2001, 2016), but not all professions have the same characteristics or can support use of large-scale studies.

This research investigated whether a content framework developed using a small-scale study that relies on the work of a carefully selected panel of experts provides adequate content-related validity evidence. Such evidence is essential to a high quality test that aligns with industry standards (AERA, APA, & NCME, 2014). The method selected was a three-phase process, referred to as a practice analysis study: 1) planning, 2) in-person meeting to develop the content framework, and 3) survey. Thirteen early childhood Montessori teachers, teacher trainers, and administrators in South Carolina were carefully selected to represent the profession in that state. These experts identified essential tasks of an entry-level teacher. A survey was then administered to both the panel and a broader sample of professionals in South Carolina to determine the importance of each task in the content framework.

Results from the panel of experts were compared to those from the broader professional population using parametric and non-parametric statistical analysis. In addition, results were evaluated within a validity framework to determine the impact when relying on a small-scale

study. Results of non-parametric analyses indicated agreement between the two groups for all but 2 of 19 tasks. Results of correlation analysis indicated overall moderate to strong agreement between groups. Selecting a small-scale study method using a panel-only practice analysis is supported based on the results, assuming the panel can be carefully selected to represent the diversity of the profession and stakes are low to moderate. There may be some concerns for using panel results alone when stakes are high, such as those for licensure testing.

ACKNOWLEDGMENTS

As with any formidable task, many individuals must come together for success. I was but one of many people on this journey, thus, there is an extensive list of those to whom I wish to express an immense debt of gratitude.

First, to my love and sugar daddy, Steve Talley; no gifts of diamonds or pearls, yachts or fancy cars could be more valuable to me than the educational pursuits you have supported during our years together. More than just tuition, you gave me the gift of freedom to pursue my passion, exercised patience and tolerance during countless hours of studying and stress, and avowed enduring love even in the hardest moments. Thank you, sweetheart.

One should never underestimate the influence that children have on our lives or how much their love and very presence can motivate one to do great things. I was, at every moment, aware of my children's watchful eyes waiting to see if I would achieve my goal. I hope what they witnessed was the importance of education and of pursuing one's passions; observing that these things come only with hard work and sacrifice. Noah and Maia, my little Montessorians, are bright shining stars in my life, giving me strength, courage, and inspiration every day of my life.

Mom and Dad, your lifelong dedication to education was an inspiration to me. Thank you for your unconditional love and support, and for helping to raise Noah and Maia. I couldn't have made it through these four years without you.

Thank you to all the friends and family, who shouted encouragements from the very beginning, tolerated my endless stories of coursework and research, shared many glasses of wine, and acknowledged the hard work and accomplishment.

To Jim Penny, I will always be grateful for your mentorship and the truly unique letter of recommendation for graduate school. It will hang on my wall next to the diploma to remind me of how far I've come with the help of others.

Stephen Johnson, my great mentor, you came to my rescue. I will always appreciate you stepping in last minute to help me finish comps and joining my dissertation committee. Thank you for taking time from the very important work you are doing now to help me achieve my goal.

Rita O'Sullivan, my wonderful advisor, I am so grateful to have had the privilege of learning from you across two graduate degrees. Your guidance throughout the PhD process, keeping me focused on the practical when my mind was wandering to the improbable, is greatly appreciated.

A graduate committee has such an onerous task, investing time, energy, and commitment to someone else's passions and goals. Thank you to Gregory Cizek for supporting me through yet another degree. I will always be grateful for your guidance as my advisor during my Master's program, continuing support me through the doctoral coursework and comps.

To Chad Buckendahl, I am so grateful you set aside time to provide guidance, answer endless questions, and attend meetings that I know were difficult given your role in our industry. I'll be a better psychometrian and researcher because of it. Thank you!

Bill Ware, you are such an outstanding teacher and I am only sorry I couldn't take more statistics courses during my doctorate program. I am so grateful to have had you as a professor and on both my Master's and Doctoral committees. Thank you for your time and support!

Sherrick Hughes, I am a more socially conscious and reflective person at the end of this process because of you. I share my heartfelt thanks for teaching me so much in one semester and agreeing to serve on my committee.

Many graduate students and professors say what a lonely rode dissertation research and writing can be. I was fortunate that I walked very little of this road alone, thanks to the partnership I established with a very talented and dedicated Montessorian, Christine Lowry. Christine, I can't imagine what this experience would have been without. Your passion for Montessori education and dedication to gaining recognition for teachers in this community was a constant reminder of how important this project was. I can't wait to see what happens next!

Thank you Paula Massengill for all the hours spent editing, making me a better writer in the process. You were a great support for me at the end of this process when I was feeling exhausted and out of patience. Thank you so much for your hard work and support!

John and Anne Moncure, I will never be able to repay your generosity in welcoming me and members of the research panel into your home and sharing with us such an exquisite dining experience! Thank you, John, for also sharing your school for this project. I am so grateful.

Finally, I thank the South Carolina Montessori Alliance, all the amazing Montessori teachers, teacher educators, and administrators in South Carolina who participated in this study. I was so generously accepted within this community; with many sharing my passion for the research and desire to share Montessori education with as many children as possible.

PREFACE

Applied measurement in education and credentialing is typically seen as a scientific endeavor, focusing on mathematics and statistics. The artistic elements of practice are often obscured. Testing professionals such as psychometricians do not typically reveal their personal biases or positionality in their work as there is a constant striving to reduce such influences. However, there is a deeply human aspect to the work of a psychometrician.

Consider the individuals taking tests, such as students and professionals seeking credentials, whose lives are impacted by the testing experience and consequences of results. The entirety of society is impacted by having an educated citizenry as evidenced, in part, by results of assessments, and credentialed professionals providing services. Many experts outside of the testing profession are involved in creating of any given assessment. There may be dozens, hundreds, or even thousands of people involved in developing any educational or credentialing test. In most cases, these individuals are not paid; they are volunteers. They come to meetings, attend web-conferences, write and review test content, aid in making critical decisions such as content standards and cut scores, and serve on governing boards.

This study was motivated by the human factor of testing. As a researcher and a parent of children growing up with a Montessori education, I have observed wonderfully talented teachers who have dedicated their lives to a concept of education that is out of the main stream and not generally accepted by the public. They work hard toward training, only to be denied acknowledgement by state departments of education. And without an industry standard for the profession, schools and children are not assured minimally competent teachers. Developing a

credentialing program and tests for Montessori teachers is going to be a challenging task, given the unique characteristics of the profession and educational community in which they work.

TABLE OF CONTENTS

LIST OF TABLES	xiii
LIST OF FIGURES	xv
CHAPTER 1. INTRODUCTION	1
Credentialing and Practice Analyses	3
Research Questions	8
Theoretical Framework	9
CHAPTER 2. LITERATURE REVIEW	24
Standards	24
Legal Defensibility	29
Review of Practice Analysis Methodology	33
Montessori Education	49
CHAPTER 3. METHODS	56
Participants	56
Phase I	62
Phase II	62
Panel meeting	63
Development of tasks	63
Phase III	65
Pilot test	67
Communication and administration of final questionnaire	68
Method for creating a test blueprint	70

CHAPTER 4. RESULTS	74
Results of Phases I and II.....	74
Phase III: Survey Results	78
Response rate.	79
Initial screening of survey responses.	80
Demographics.	83
Analyzing individual tasks.....	100
Analyzing the measurement scales using RRSM.	103
Analyzing combined scales.....	106
Group comparisons.	110
Additional tasks from Survey participants.....	120
Summary	121
CHAPTER 5. DISCUSSION	122
Evidence Based on Representation.....	122
Question 1: Final ratings.....	124
Question 2: Group comparisons.....	125
Question 3: Sufficient evidence for validity	127
A second interpretation of results	132
Limitations of the Study.....	133
Implications for Current and Future Research.....	134
APPENDIX A. SUMMARY OF PUBLISHED PRACTICE ANALYSIS STUDIES	138
APPENDIX B. LIST OF PARTICIPATING SCHOOLS	143
APPENDIX C. DESCRIPTION OF STUDY FOR POTENTIAL PARTICIPANTS	145
APPENDIX D. CONSENT FORM FOR ADVISORY GROUP	149

APPENDIX E. PHASE I. ESTABLISHING FOUNDATIONAL INFORMATION FOR THE PRACTICE ANALYSIS PANEL	154
APPENDIX F. PANEL MEMBER CONSENT FORM.....	172
APPENDIX G. MEETING WORKBOOK	177
APPENDIX H. INITIAL TASKS DEVELOPED IN SMALL GROUPS	189
APPENDIX I. PANEL MEETING EVALUATION QUESTIONNAIRE	192
APPENDIX J. PILOT QUESTIONNAIRE.....	193
APPENDIX K. EMAIL COMMUNICATIONS FOR SURVEY	216
APPENDIX L. FINAL QUESTIONNAIRE	220
APPENDIX M. EMAIL INVITATION FOR FINAL SURVEY.....	246
APPENDIX N. METHODS COMPARISON FOR COMPUTING TASK WEIGHTS USING SURVEY RESPONSES	249
APPENDIX O. RELIABILITY ANALYSIS ACROSS METHODS FOR ESTIMATING TASK WEIGHTS.....	250
APPENDIX P. BLUEPRINT WEIGHT COMPARISON	251
APPENDIX Q. FREQUENCY TABLES WITH PANEL AND SURVEY GROUP RESPONSES TO CRITICALITY AND FREQUENCY SCALES FOR EACH TASK	252
APPENDIX R. FREQUENCY TABLES WITH PANEL AND SURVEY GROUP RESPONSES TO COMBINED SCALE FOR EACH TASK.....	271
APPENDIX S. GRAPHS COMPARING MEAN RANK IN MANN-WHITNEY U ANALYSIS.....	290
APPENDIX T. ADDITIONAL SUGGESTIONS FROM SRVEY PARTICIPANTS FOR EACH CONTENT DOMAIN.....	292
APPENDIX U. SUMMARY OF COMMENTS UNDER EACH DOMAIN IN THE SURVEY.....	296
REFERENCES	299

LIST OF TABLES

Table 1.1. Example of a Task Statement with Associated Knowledge and Skills	4
Table 2.1. Summary of Content-Related Standards for Educational and Psychological Testing	26
Table 2.2. Summary of Content-Related NCCA Standards	27
Table 2.3. Summary of Content-Related ANSI Standards	27
Table 2.4. Summary of Content-Related EEOC Statutes	28
Table 2.5. MACTE Accreditation Affiliates.....	52
Table 3.1. Summary of Sampling by Schools and ECMTS	58
Table 3.2. Recruitment Plan for Panel SMEs	61
Table 3.3. Summary of Practice Analysis Meeting Feedback	65
Table 3.4. Rating Scale Options Presented to SME Panel.....	67
Table 3.5. Summary of Pilot Evaluation Questions.....	68
Table 4.1. Content Domains and Tasks	76
Table 4.2. Summary of Responses to the Survey	79
Table 4.3. Summary of Total Scores for Panel and Survey Groups	81
Table 4.4. Comparison of General Demographics between Panel and Survey Respondents	85
Table 4.5. Comparison of Geographic and Population Regions between Panel and Survey Respondents	87
Table 4.6. ECMT Comparison of Education and Experience	89
Table 4.7. Additional Roles of the ECMTs	90
Table 4.8. Summary of Education Affiliation of ECMTs	91
Table 4.9. Comparison of Professional Certifications and Practice Settings	93
Table 4.10. Comparison of Teacher Trainers' Level of Education, Educational Affiliation, Years of Experience, and Credentials	95
Table 4.11. Affiliations Represented by Teacher Trainers	97

Table 4.12. Comparison of Education and Experience of Administrators of Montessori Programs	98
Table 4.13. Comparison of Montessori Certification and Practice Setting in Which Administrators Work.....	99
Table 4.14. Analysis of Criticality Responses	101
Table 4.15. Analysis of Frequency Responses	102
Table 4.16. Reliability of Criticality and Frequency Scales	106
Table 4.17. Analysis of Combined Responses.....	107
Table 4.18. Reliability Comparison	109
Table 4.19. Summary Mann-Whitney U test of Rank Order Between Independent Groups	113
Table 4.20. Comparison Between Groups of Proportion of Test Content.....	116
Table 4.21. Summary of Correlation Analysis for All Three Response Scales.....	119

LIST OF FIGURES

Figure 1.1 Validation process based on Kane’s argument-based approach.....	15
Figure 1.2. Argument-based approach as a theoretical framework for establishing content-related validity evidence.	22
Figure 2.1. Summary of path to certification for Montessori teachers.	52
Figure 4.1. Distribution of criticality ratings for survey and panel scales.....	82
Figure 4.2. Distribution of frequency ratings for survey and panel scales.	83
Figure 4.3. Comparison of panel, survey, state ECMTs by geographical region.	87
Figure 4.4. Comparison between panel and survey participant affiliation.	92
Figure 4.5. Comparison of criticality rating scales.	104
Figure 4.6. Comparison of frequency rating scales.	105
Figure 4.7. Comparison of combined rating scales.....	108
Figure 4.8. Q-Q Plots using panel and survey <i>b</i> parameters to test for normality.....	109
Figure 4.9. Comparison between survey and panel responses across mean scores.....	114
Figure 4.10. Proportion of test comparison between panel and survey responses.	117
Figure 4.11. Comparison of mean <i>b</i> parameters for combined, criticality and frequency scales.	119

CHAPTER 1. INTRODUCTION

Consider a test intended to assess the knowledge and skills required to competently teach in a Montessori classroom. Developing such a test requires many critical decisions regarding application of measurement theory and methodology within a particular context. A delicate balance of art and science is required to meet quality standards of a testing system that incorporates program governance, policy, and the test itself (AERA, APA, & NCME, 2014; Cizek, 2001; Lane, Raymond, & Haladyna, 2016). This balance is achieved when methods supported by current research are appropriately selected and applied, taking into account contextual factors such as test purpose, characteristics and needs of the target testing population, program goals and limitations, and program and industry policies. Making appropriate selections assumes availability of current research; otherwise, methods tend to become stagnant with choices generally following a model of what has always been done as opposed to what is likely to be most effective in a given situation.

For example, determining what knowledge and skills should be included on the aforementioned test for a Montessori teacher would typically begin with selection an appropriate method for defining the content framework that incorporates a method for evaluating and quantifying that framework in order to establish a test plan or blueprint. Although there are multiple ways in which this may be accomplished, the most common method found in the literature utilizes a group or panel of subject matter experts (SMEs) in the specified profession defining the content framework, followed by a large-scale survey of the broader professional population (Raymond, 2005). The latter is often the focus of the reported study, with little

information provided on how the content framework was determined (Bradley, Bergen, Ginter, Williams, & Scalise, 2010; Darling, Fleming, & Cassidy, 2009; McKenzie, Dennis, Auld, & Lysoby, 2016; Tsai & Kramer, 2014). In a review of related methods, Raymond (2001) suggested that there is an “assumption that large-sample practice analyses are superior to committee-based practice analyses,” referring to a study that focuses on a large-scale survey rather than the work a small, carefully selected, and well-trained panel of SMEs. There is, however, little empirical evidence to support this assumption. Raymond pointed to two studies that challenged this assumption by examining the value of a survey compared to the work of the committee: Tannenbaum and Wesley (1993) and Maurer and Tross (2000). Both studies compared responses to a questionnaire derived from a survey of representative professionals to those from a carefully selected panel of experts. Researchers concluded that opinions between groups were similar enough to support use of the panel results alone to determine the content domain of a credentialing test. These studies suggest that a large-scale survey should not be the default method for determining content of a credentialing test.

This line of inquiry ended here (Maurer & Tross, 2000), with limited insight into the viability of using a panel-only approach or some variation that excludes the use of a large-scale survey. Without current empirical studies that support the use of alternatives to large-scale survey methods in determining test content under different conditions, justifiable options are limited, particularly for small-scale testing situations where large samples of professionals are not available. The purpose of the current study was to address this lack of research, using the example of a Montessori teacher test of knowledge and skills.

This study focused on an inquiry into panel-based practice analysis methods within a measurement validity framework, using a new credentialing test for early childhood Montessori

teachers as the subject of the study. The following sections provide an introduction to the theoretical framework, credentialing and practice analysis methodology, and specific research questions to be answered by this study.

Credentialing and Practice Analyses

Kane (1994b) stated, “the more complex, elegant, and sophisticated our methods, the easier it is to get engrossed in the details of the methodology and forget the original question” (p. 134). The original question in developing a test is whether choice of methods produces outcomes that support the established purpose of the test as well as intended interpretations and uses of test scores. It is about validity. The Standards for Educational and Psychological Testing, to be referred to herein as the Standards (AERA et al., 2014), define test validity as “the degree to which evidence and theory support the interpretations of test scores entailed by proposed uses of tests” (p. 9). In terms of application, the definition implies that methods selected and decisions made in the development of a test carry a heavy burden and must be justified, both theoretically and empirically.

Establishing validity for a certification or licensure test, together referred to as credentialing (AERA et al., 2014; Knapp & Knapp, 1995; Raymond & Neustal, 2006), typically begins with a practice analysis study. A practice analysis is a systematic method for determining the professional competencies that are relevant for granting credentials and establishing content-related evidence in support of interpretation and uses of test scores. The authors of the Standards have acknowledged the primacy of such evidence for a credentialing test and indicated the essentiality of using appropriate methods that clearly link the content of the test to the professional domain of practice (Standards 11.2, 11.3).

In a practice analysis, competencies are typically first delineated as critical tasks that are performed on the job; followed by knowledge, skills, and abilities (KSAs) necessary to perform these tasks. Tasks and KSAs are often grouped into meaningful categories, or domains. Table 1.1 provides a hypothetical example of a task with associated knowledge and skills for the role of a Montessori teacher.

Table 1.1. Example of a Task Statement with Associated Knowledge and Skills

Task	Evaluate mastery level of individual students using formative assessment techniques in order to provide differentiated instruction.
Knowledge	Indicators of different mastery levels for a given subject or work. Formative assessment techniques.
Skills	Applying formative assessment techniques to evaluate mastery. Determining specific instructional need of students based on observations.

Many terms are used to describe this methodology including job analysis, job task analysis, role delineation study, and occupational analysis (Kane, 1997; Nelson, 1994). Kane (1997) provided a clarification of the key difference between a practice analysis and a job analysis stating that “a job analysis suggests that the purpose of the investigation is to identify the requirements of a specific job. Credentialing tests incorporate competencies across a wider range of settings that are not ordinarily job specific” (p. 6). For example, a teacher may work in a public or private school, and in a rural or urban setting. Based on this description, the term practice analysis is adopted for this study, but with a point of clarification.

Although there are numerous methods described in the literature for conducting a practice analysis, the primary method used for credentialing tests is to first assemble a representative

committee or panel of SMEs in order to define the relevant tasks necessary for competent performance in a clearly defined profession (Knapp & Knapp, 1995; Raymond, 2005; Raymond & Neustal, 2006). This method aligns with the expectations of the Standards (AERA et al., 2014), that the content is based on judgements by experts. KSAs may also be developed by this panel or may be developed separately at a later time by this same group of SMEs or a different group. The SME panel may be provided with some content to start from, including the work completed by a previous analysis or an initial attempt at defining the content domain by a testing professional, job analyst, and/or a different panel of SMEs. For example, the present study will use a small advisory group of representative experts to establish foundational information in support of the work to be conducted by the larger SME panel.

In the testing profession, work accomplished by the SME panel is typically followed by surveying a representative sample of professionals on the importance of the established tasks (Raymond, 2005; Raymond & Neustal, 2006). There are two goals for this step: (1) verifying that the work of the panel represents the view from the broader profession, and (2) to quantify the importance of each task as it relates to the profession and apply those quantifications (i.e., ratings) to establish a blueprint or test plan. Indication of importance, as required by standards (AERA, APA, & NCME, 2014; International Standard Organization Committee on Conformity Assessment, 2012; National Commission for Certifying Agencies, 2016) is required in various standards is measured using various Likert-type response formats, referred to herein as rating scales. In the blueprint design phase, high overall ratings would result in more items assigned to a given task on the blueprint, and low ratings would indicate fewer. The term item is used in this study to denote all types of prompts to which a test taker responds. This description is an

oversimplification of developing a blueprint, but the subject will be revisited again in the literature review as well as the description of study method and results.

The use of a survey is not specified in the Standards (AERA et al., 2014), but is what is used in most practice analysis studies for the purpose of including a broader sample from the different practice settings and demographics (Maurer & Tross, 2000; Raymond, 2005, 2016; Tannenbaum & Wesley, 1993). In a review of relevant research, few studies were found that investigated the value of a survey for verifying the work of a SME panel, particularly when the testing population is very small.

Current methods originate from the field of Industrial and Organizational (I/O) Psychology, describing methods created for human resource purposes such as personnel selection and job placement (Gael, 1983, 1988b, 1988c). Most of the cited texts in practice analysis studies published in the past 10 years date back to the 1980s with supporting research for the methodologies going back to the 1970s. Thus, there is little current research to support the claim that surveys provide additional information or adequately verify the work of the panel. In fact, some scholars questioned the value of the survey portion of the study, particularly for complex professions where the typical questionnaire cannot represent such complexity (Kane, 1997; Raymond, 2005). Although there are other ways in which a test blueprint can be established, including using SME judgments (Maurer & Tross, 2000; Tannenbaum & Wesley, 1993), most testing programs reviewed in the literature indicated the use of a questionnaire, emphasizing this phase of the study over the contribution of the panel. These studies focused on the content-related outcomes of the practice analyses rather than the methodology, citing methodological reviews such as Raymond and Neustal (2006) and Knapp and Knapp (1995), but no empirical research in support of methodological choices.

Two published studies, Tannenbaum and Wesley (1993) and Maurer and Tross (2000), were found that directly related to the use of surveys in comparison with the SME panels for determining the content domain. The stated purposes of these studies indicated the potential burden of large-scale surveys in a practice analysis, citing issues of time and expense, and inquiring whether such a burden was justified. In both studies, researchers compared SME panel ratings of job task importance to the ratings from a sample of professionals responding to the same questionnaire. Researchers in both studies drew similar conclusions indicating that panel and field respondent ratings were very similar and that the surveys may not significantly add to the final results of the practice analysis. There are limitations in the generalizability of these studies that will be described in detail in the review of literature. No studies beyond these two were identified that further investigated these findings.

Although these studies make a salient point about the demands of the survey portion of a practice analysis on a testing program's resources, it may be even more important to consider the limited utility of survey results in defining a test blueprint when the population or available sample from the population is limited. In such a situation, the results of the survey may not warrant the emphasis that test professionals tend to place on them. Instead, the opinions of the panel may be sufficient to support test inferences from a content perspective.

Consider the example of a new testing program in an industry with a small number of professionals such as early childhood Montessori teachers. The field of Montessori education currently has an established credential awarded after completion of a teacher-training program. It is important to note the distinction here between an assessment-based certificate program and a certification program for the purposes of this study.

An assessment-based certificate program is one in which the training and the assessment are linked (Fein, 2012). The candidate must complete training and then pass a test, sometimes after a certain period of time (i.e., one month). A certification program requires candidates to pass a test or set of tests. There are typically a set of eligibility requirements that must be met in order to qualify to take the test. For example, test takers may be required to hold a bachelor's degree in a particular field from an accredited institution and have one year of work experience. The test, however, is not aligned with a single training program or curriculum.

Although one is not preferable, in general, over the other, Fein (2012) pointed out that there is a “desirable level of objectivity to have the assessing entity be independent of the training provider” (p. 15). An important distinction is one of test purpose and the claims that may be made based on test scores and supported by the available validity evidence. Claims of competency beyond what is included in the training may not be made for a certificate program where the test is constructed to align specifically with that training. A certification test, however, aligns with the tasks and KSAs that are critical to competent performance on the job, thus, more broad claims about competency may be made.

Research Questions

The proposed research questions are focused on the methodology used to define the practice in order to establish the potential credentialing test content domain for early childhood Montessori teachers. The study conducted to address these questions was a three-phase practice analysis, similar to those described in the literature. The first two phases involved initial planning work by an advisory group (AG) of SMEs followed by a SME panel who were asked to define the critical tasks necessary for an entry-level early childhood Montessori teacher. The final phase

of the study was the development and administration of a questionnaire, and administered to a group of representative early childhood Montessori teachers, teacher trainers, and administrators.

The study was designed to answer a single overarching research question and three sub-questions:

- Are the resulting task ratings determined by a carefully selected panel of subject matter experts and a sample of subject matter experts surveyed from the broader target population substantially different, such that the absence of a survey would negatively impact the strength of the validity argument for a credentialing test?
 1. What are the final task ratings and blueprint weights based on responses to a questionnaire by the selected panel of subject matter experts?
 2. How do the final task ratings and blueprint weights generated from surveying representative professionals compare to those of the panel?
 3. How do these results provide content-related evidence for the validity argument in support of test scores that would be generated from a future certification test developed based on this study?

The study was conducted within Kane's (1994a, 2006, 2013) validity framework. This framework provided a theoretical justification and structure for the methodology. The argument-based approach is described next.

Theoretical Framework

Validity as a theoretical framework for this study begins with what Kane (1997) referred to a "chain of inference," which describes how scores are interpreted and the explicit claims made based on those interpretations. For example, envision a hypothetical test for an entry-level Montessori teacher. An individual takes a test and receives an evaluation of performance in the

form of a score or a pass/fail decision. The outcome (score or pass/fail decision) is interpreted to mean that the test taker has or has not demonstrated competency in the specified content domain. From that interpretation a claim is made such that, based on the results, the test taker is competent, or not, to practice as an entry-level Montessori teacher.

In order to justify this chain of inference there must be supporting evidence. For example, the inference that a test taker who has achieved a passing score has demonstrated competency in the specified content domain requires some evidence that the test reflects the specified content domain, and that the scores fairly and reliably evaluate performance in the domain. The claim that the test taker is competent to practice in a profession requires evidence that the content is related to the profession.

Collecting evidence of score interpretations and use is a complex endeavor. The Standards (AERA et al., 2014) provide guidance for what constitutes adequate evidence, but not a precise framework or instructions for how to collect and present this evidence. Kane (1992, 1994a, 2006, 2009) proposed a practical model for supporting the process of validating test scores, referred to as the argument-based approach. The goal of this approach is to provide a framework for collecting and synthesizing evidence that can withstand challenges. A more refined examination of the current view of validity and the types of validity evidence provide the foundation of this framework.

Evidence of validity. What was once a fragmented framework made up of multiple types of validity is now a single, unitary view of validity; multiple types of validity evidence are necessary to support the interpretations related to the desired construct (Cizek, 2012; Kane, 2006; Messick, 1989). This view emerged over several decades, beginning with the introduction of construct validity by Cronbach and Meehl (1955) and the eventual elevation of construct

validity from one type of validity to the whole of validity (1989). With this came an obligation placed upon test developers to define and present evidence that a test is measuring the desired construct via multiple indicators of the construct laid out as a nomological network. Content and criterion-related evidence constituted that network of indicators. This approach caused some angst among practitioners (Borsboom, Mellenbergh, & van Heerden, 2004; Lissitz & Samuelsen, 2007; Sireci, 2007) attempting to operationalize this theory, particularly with regards to credentialing tests where criterion-related evidence is difficult to produce due to a lack of comparable measurements to serve as a criterion (Kane, 1994a). Kane (2013) addressed this issue with the argument-based approach by replacing the theoretical network with a network of interpretation-related assumptions and evidence in support of those assumptions.

The current Standards (AERA et al., 2014) exemplify the unified concept of validity, but without the focus on construct validity per se, defining construct as “the concept or characteristic that a test is designed to measure” (p. 11). Evidence is then presented in support of the designated score interpretation for the specified use rather than on the development of complex theoretical networks except in cases where such complex theory is warranted. There are five primary types of evidence discussed in the Standards: content, response processes, internal structure, relation to other variables, and consequences of testing. Although the Standards provide some guidance as to the sources of information and types of studies that might be considered for each type of evidence, it is not intended to be prescriptive and additional thought must be given to what constitutes evidence in a particular situation.

For credentialing tests, content-related evidence is a primary source of evidence. It supports the link between what is intended to be measured - the construct- and score interpretations and their intended use (AERA et al., 2014). Since most constructs would be too

broad to measure all related content in a single test, choices must be made regarding which content best represents the domain given the purpose of the test (Buckendahl & Plake, 2006; Kane, 2006; Sireci, 2007). These choices require justification through empirical and analytical evidence, although content-related evidence is primarily analytical. Content-related evidence may include, but is not limited to, documentation of the development and review of test specifications or blueprint (practice analysis), item development and review process, alignment of items with the blueprint, and development of performance level descriptors (Buckendahl & Plake, 2006; Schafer, Wang, & Wang, 2009).

Most of the evidence in this category is based on professional judgement. For example, items are written and reviewed by SMEs. Alignment studies require SMEs to make value judgements regarding alignment of items with the test blueprint (Li & Sireci, 2013; Sireci, 2007). Standard 1.9 (AERA et al., 2014) indicates that information regarding the experts involved in the test development processes and decision making should be documented. This standard is applicable across all of the types of evidence whenever expert judgment is used in test development and in establishing evidence of validity.

In addition to evidence supporting the content selected for tests, consideration should also be given to what was not selected to be on the test, and the rationale for the choices in relation to the specified interpretation (Kane, 2006, 2013). Evidence of this type is valuable in addressing challenges to validity or “rival hypotheses” (AERA et al., 2014) based on construct underrepresentation, which is a source of validity error that arises when some important aspect of a construct is not included on a test. This evidence may be in the form a narrative describing the process that was used with the practice analysis panel to determine what tasks were included in

the final content domain and tasks or domains that were eliminated (possibly after the survey was administered).

Validation must also be conducted within the limitations of the testing program or system (Jones, Smith, & Talley, 2006). Choices have to be made about which information provides the strongest body of evidence in support of score interpretations and use. This challenge demands a guiding framework to structure such evidence in a comprehensive manner. The argument-based approach as described by Kane (2006, 2013) provides such a framework to evaluate and prioritize the available evidence and integrate that evidence to make a case for validity of score inferences.

The argument-based approach. In an argument-based approach the interpretation of scores for a particular use, related assumptions and claims, and supporting evidence are articulated in the form of two types of arguments (Kane, 1992, 2006, 2013). The first argument is a statement referred to as an interpretive argument, which takes the form of an argument in that it has a premise (test scores) on which conclusions are based (interpretation of or claim based on test scores). For example, in the case of the Montessori teacher certification test referred to in the introduction, the interpretive argument might state that a passing score on the test (premise) may be interpreted to mean that the test taker has the minimum knowledge and skills necessary to practice as an entry-level teacher in a Montessori classroom (the claim). There are a number of possible intended uses for this certification, including use as a voluntary certification or as fulfilling a requirement for licensure.

Interpretive arguments should be deliberate and well thought out from the beginning of the test development process, when possible (Kane, 2006). This does not imply that the argument cannot be changed or that new interpretations and uses are not added. However, the interpretive

argument is intended to be a dynamic statement that evolves as the test is developed and new evidence accumulates throughout the development and evaluation process. New counterarguments or alternate assumptions may emerge that require adjustments to the argument.

The purpose of the second argument, the validity argument, is to establish plausibility of the claims made based on explicit and implicit assumptions of the interpretation and use. These assumptions will be described in further detail in a moment. Unlike the interpretive argument, it is not simply a statement or series of statements, but a body of evidence. Kane stated that “the validity argument provides an evaluation of the interpretive argument. To claim that a proposed interpretation or use is valid is to claim that the interpretive argument is coherent, that its inferences are reasonable, and that its assumptions are plausible” (2006, p. 23). Plausibility is substantiated by evaluating the strength of the available evidence in support of the interpretation and use. In the example of the teacher certification, the plausibility of the claim that a passing score indicates competency as an entry-level teacher will be established, at least in part, by the practice analysis results that define what tasks an entry-level teacher must be able to do and the knowledge and skills necessary to perform those tasks. This is one source of evidence, which is content-related.

The validity argument is also dynamic; a potentially ongoing, iterative process of evaluation, but not one without end (Kane, 2013). Evaluation of the interpretive argument begins with assessing the clarity and cohesiveness of the argument including the claims being made and the specified use.

Implementation of the argument-based approach. Kane (2006, 2013) divided the validation process into two stages, as shown in Figure 1. Initial supporting evidence for a validity argument is collected within the Design Stage of validation and begins with defining the

purpose, intended interpretation, and use of test results (interpretive argument). Also in the Design Stage is defining a content domain for the test and developing the test itself (Kane, 2006). For example, content-related evidence collected through a practice analysis for the development of a credentialing test is collected in the design stage. Kane maintained that the interpretive argument, more specifically the intended interpretation and use of scores, should drive test development. Although there are instances where validation does not occur simultaneous to the development of the test, this approach emphasized the scenario where it does and the advantages therein.

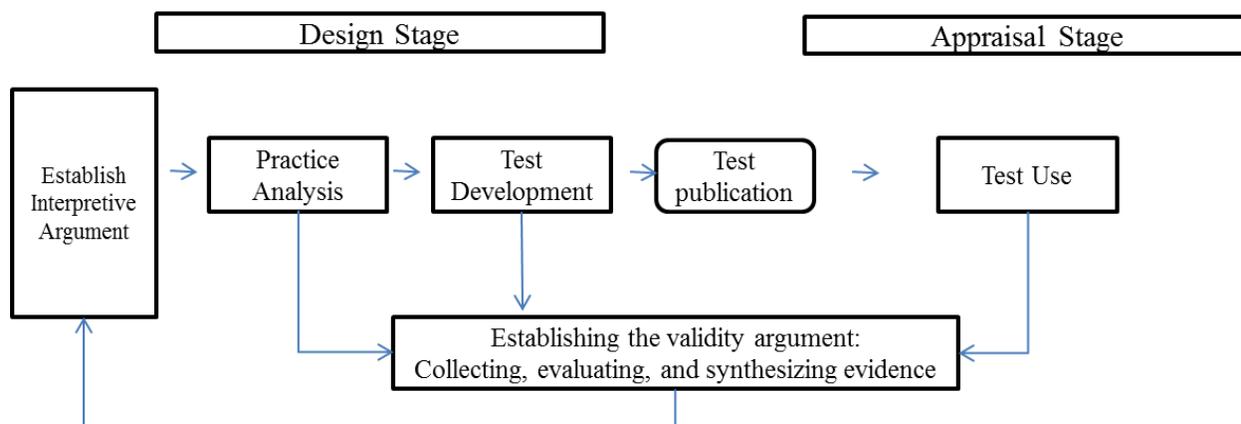


Figure 1.1 Validation process based on Kane’s argument-based approach.

In the Appraisal Stage, which occurs after the test is operational, additional evidence may be collected through validity studies such as comparisons of the test with other related measures (criterion studies) and predictive studies that indicate how well scores on the test predict actual performance (Kane, 2006). Additional analysis of operational scores in the Appraisal Stage may be useful for conducting supplemental evaluation on the functioning of the test, particularly when adequate samples of test takers were not available for field testing. During this stage the

evaluation approach shifts from being confirmatory (the focus of the design stage) to critical. Appraisal requires a more external, judgmental view, focusing on the weakest assumptions underlying the interpretive argument and investigating possible counterarguments.

Kane (2006) defined a final component of the argument-based approach based on Toulman's (1958) approach to practical arguments, which is intended to provide a chain of reasoning from claims made based on observations or data to the evidence in support of the claims. This expands Kane's framework to include warrants for each claim, which are general rules or laws that support the inference from an observed test score to performance, and the necessary backing for those warrants (Kane, 2006, 2013). A warrant related to content evidence would be based on the methodology of a practice analysis study. Warrants give justification for linking observed behavior on a test to the actual performance domain. In an example where the claim is strong and states that an observed score indicates competent job performance, a predictive validity study may be conducted to support this link whereby the warrant is the regression analysis used to demonstrate this link.

In credentialing, programs typically do not make such strong claims because establishing the link is difficult, thus a weaker claim is stated (Kane, 2013). In credentialing, the claim typically states that a candidate has the minimum knowledge and skills necessary for competent performance, not that the candidate will actually perform well. For example, a claim based on a passing score of a teacher certification test may be that the candidate has mastered the knowledge and skills necessary to be minimally competent in the classroom. This claim assumes that test content aligns with the knowledge and skills necessary for competent performance and the warrant is a process as opposed to a rule in this case: the method used to define the test content domain. The warrant is backed by the resulting test plan or blueprint and, later in the test

development process, verification of alignment between the blueprint and actual test content. In this way, plausibility of claims and their assumptions are evaluated using the validity argument.

Observable attributes. Kane (2006, 2013) presented an additional layer to the framework by discussing interpretations of scores in terms of what explains test scores. At the most fundamental level is an observable attribute, which Kane refers to as, “how well a test taker will perform on average over a target domain of possible observations” (2006, p. 22). For an interpretation of scores based on an observable attribute, a target domain must be specified (i.e., teaching skills). The target domain encompasses all KSAs one could need in order to be competent with regards to the attribute in addition to the possible contexts in which these tasks may be performed (i.e., practice settings or types of classrooms). For a Montessori teacher certification, the target domain would be all subject matter and pedagogy required to competently teach in a Montessori classroom.

There is, however, only a subset of the target domain that can be measured in a standardized manner given the constraints of a testing program. To test the entire domain would likely require hundreds of test items in multiple item formats and hours or days of test taking time. Many of the areas may only be viable under rare or very occasional situations, such as working in a small rural school versus working in a larger suburban school. Thus, the target domain is reduced to the more constrained universe of generalization, or a subset of content, that will be sampled for inclusion on the actual test (Kane, 2006, 2013). Generalization from an observed score, or the actual test score achieved by the test taker, to a universe score, the score that is assumed that test taker would achieve on any number of parallel tests given under identical conditions, assumes that the test taker would perform the same regardless of which content and items were drawn from the universe of generalization. This implies consistency not

only across items, but across test forms and any other facets that might be applicable (i.e., raters, test administration conditions). The inferences made from the observed test scores generalize to a hypothetical score from the universe of generalization and then to the target domain, which represents the actual observable attributes. Thus, an additional assumption of test score interpretations is that the content of the universe of generalization is representative of the actual KSAs in that target domain. For a credentialing test, this assumption is satisfied by conducting a practice analysis study. In a typical practice analysis, SMEs are tasked with selecting the most critical competencies of the target domain to be included in the universe of generalization, which will then be used to develop the test specifications or blueprint.

Interpretive argument for a credentialing test. Consider an example of this concept in the context of a new certification for a Montessori teacher. A target domain exists with all possible tasks that an entry level Montessori teacher needs in order to competently teach across a variety Montessori classrooms. A subset of the most critical and measurable tasks from that domain is identified to be included on the test, which is the universe of generalization. If a subsequent test contains a sufficiently large, representative, and quality sample of items from the universe of generalization, then the observed score can be generalized to the universe score and finally to the target score. Three types of inferences supporting the interpretation of scores are pivotal to supporting this aim and rely heavily on content-related evidence (Kane, 2006, 2013): scoring, generalization, and extrapolation.

Scoring. The scoring inference establishes the scoring rule and its rationale (Kane, 2006). For example, it may be determined that a multiple choice test with 50 items will be scored using classical number correct scoring followed by scores rounded to the nearest whole number. A salient point in this example is that the interpretation of the score should be aligned with the

final, reported score (Brennan, 2013), which is what the stakeholders interpret and use. Thus, part of the evidence in support of accuracy of score interpretations is not only the scoring rule, scoring process and rationale for the choice of rule (i.e., choices regarding number of points given per item), but anything score-related that impacts the interpretation. This includes evidence related to test administration, score reports, and materials provided to test users and takers that describe the meaning of scores. It may also include content-related evidence, as described earlier, particularly for more complex scoring models that make use of sub-scores. When reporting sub-scores, content-related evidence may be used as justification for the domain structure that determines how items are grouped for scoring. In the case of performance tests, the scoring protocol, rubrics, qualifications and training of judges, and quality assurance procedures would also be relevant evidence under the scoring inference (AERA et al., 2014; Kane, 2006).

Generalization. The second type of inference is generalization. As specified earlier, evidence is needed to justify generalization from the observed score to a universe score (Kane, 2006, 2009). The universe score is the hypothetical score that a test taker would achieve if different samples of content were selected repeatedly from the universe of generalization; without retaining any knowledge from one administration of the test to the next. It is essentially assuming that the observed score would be the same regardless of which sample of content and tasks were presented, realizing some amount of measurement error will result. To make such a strong assumption requires evidence that the content on which scores are based (content and tasks presented on the test) is representative of the content and possible tasks in the universe of generalization. The backing for this warrant relates to the sampling on which the empirical analyses of test items and forms were conducted. Samples must be sufficiently large and representative to support analysis of items and forms.

The types of evidence are, thus, primarily those related to content representation and internal structure of the test including the choice of measurement model and reliability analyses. Evidence would include content-related evidence (practice analysis and development of the test specifications or blueprint, item selection across forms, content equivalency of forms), reliability analyses (internal consistency, generalizability studies), documentation of measurement error, and descriptions of the samples used to do analyses including consideration of representativeness and size of samples (Kane, 2006). Kane (2013) indicated the need for evidence “that the sampling was consistent with the statistical model being employed and on generalizability (or reliability) analyses (or item response theory [IRT]-based analyses) indicating that the sample was large enough to control for sampling errors” (Kane, 2013, p. 14). Evidence in support of this inference also pertains to the inference of extrapolation.

Extrapolation. Extrapolation represents a third type of inference, which makes the leap from the universe score as represented by the observed score to the actual behavior or skill defined by the attribute (Kane, 2006). Consider the chain of inference presented in Figure 1. In order to support the links in the chain that a passing score on a test indicates mastery, which then implies competency in the profession, requires extrapolation from that observed score to the actual performance domain (i.e., application of knowledge and skills in the role of a teacher). In other words, the assumption is that success on the test is reflective of competency on the job.

The further the content and response processes required for test performance are from the actual tasks the teacher performs in the classroom, the more difficult it will be to extrapolate from the observed score to target domain. A performance test such as a flight test to license an airline pilot would not require evidence of extrapolation since it is directly measuring the attribute of interest. If, however, such skills were measured on a test of selected response items,

then target scores must be extrapolated from the observed scores, thus requiring some evidence that extrapolation is warranted. In other words, there would need to be supporting evidence that test takers who pass the written test can competently fly a plane.

Evidence supporting this warrant would provide backing for the relationship between the test items and skills in flying. This evidence could include content-related evidence and rationale for the choice of item types to measure competency. Strong evidence would be criterion-related, demonstrating relationships between the selected response test and other measures of the skill such as a performance test. A predictive study could be conducted to demonstrate that performance on the test is predictive of performance on the skill (i.e., flying). This is an extreme example intended to illustrate a point about extrapolation. A written test alone would not be accepted in reality, but might be one of a set of tests that are required for licensure. Similarly, teacher certification typically includes a written test with selected response items that require extrapolation from performance on that test to a teacher's ability to perform in the classroom. This is in addition to completion of a teacher education course of study and practicum.

An important assumption of the extrapolation inference is that the variability in scores is due to the construct being measured and not from construct irrelevant sources of variance (Kane, 2006). Test format is a possible form of invariance that should be examined, such as the example of the selected response pilot test. Scores on the test may be due to test-wiseness or guessing rather than flying skills. One of the ways in which Kane suggested increasing ability to extrapolate and keep construct irrelevant variance at a minimum was to use multiple types of observations in a test (i.e., multiple-choice with open-ended tasks) or in the criterion measures used for comparison.

Applying the argument-based approach to the current study. Although various types of evidence are necessary in establishing validity, in credentialing, content-related evidence provides the most significant support for score interpretations and use (AERA et al., 2014; Kane, 1994a). The Standards (AERA et al., 2014) indicate the need for a strong link between the content domain of the test and the professional practice domain. The way in which these two domains are linked is through the inferences and related evidence described above.

Figure 2 presents this concept visually. At the top of the figure is valid interpretation and use of test scores. Appropriate interpretation and use of scores are supported by the three inferences just described: scoring, generalization, and extrapolation. Each of these inferences can be supported, in part, by content-related evidence. The link between content-related evidence and the scoring inference is denoted as a dashed line since this inference is only supported by content when the domain structure defined by that evidence is used in establishing scoring rules.

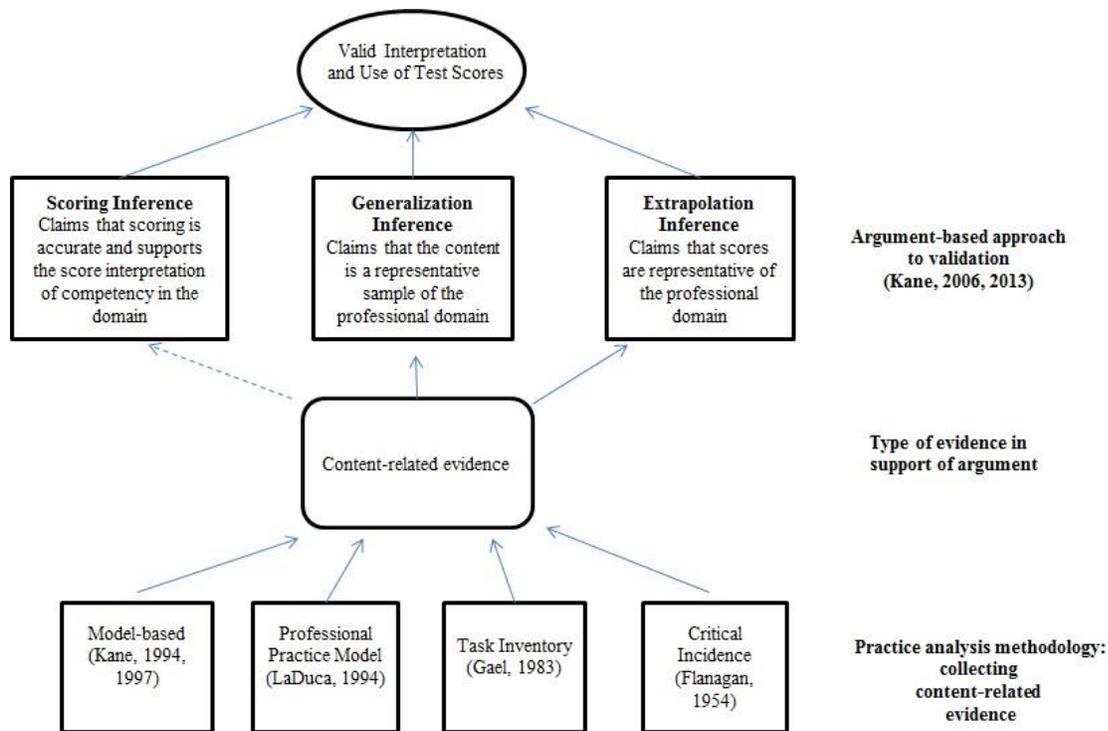


Figure 1.2. Argument-based approach as a theoretical framework for establishing content-related validity evidence.

Finally, the bottom segment of the figure provides possible practice analysis methods by which the content-related evidence may be established for a credentialing test. The four methods presented in the model are a sample of all possible methods. A description of the primary methods used to establish content-related validity evidence will be presented in the following review of literature, focusing on the most prevalent methods found in published studies in the past 10 years.

CHAPTER 2. LITERATURE REVIEW

Literature reviewed in this chapter describes commonly selected practice analysis methods as well as factors that impact selection and implementation of those methods. These are presented in three general areas: (1) test development standards, (2) legal defensibility, and (3) common practice analysis methods. Following this review is a description of the general landscape of Montessori education and current Montessori teacher training credentialing. This information provides important context for understanding the subject matter and results of the current study. The first two sections are a review of industry standards and relevant legal statutes that may influence practice analysis method selection and execution.

Standards

When selecting a practice analysis method it is important to consider how potential methods fulfill requirements of relevant standards, particularly with regards to providing necessary validity evidence. Determining whether standards are relevant to a particular test may begin with reviewing The Standards for Educational and Psychological Measurement (the Standards) (AERA et al., 2014) discussed in Chapter 1. These standards provide guidance for, rather than a predetermined set of, steps in constructing measurement instruments for educational, psychological, credentialing, personnel selection, and program evaluation purposes. In other words, they are not intended to be prescriptive, as the specific methods used in test development must be determined within the context of each situation. Although not legally binding, the Standards are a representation of cross-industry, evidence-based best practices in test development.

Other relevant standards are specific to member organizations that accredit credentialing programs: National Commission for Certifying Agencies (NCCA) and the American National Standards Institute (ANSI). These standards designate requirements for developing high quality credentialing programs, including all elements of the testing system (i.e., governance, test development, policies and procedures for administering a program). They also provide a guiding framework for evaluating the quality of a credentialing program and associated tests. The requirements specified for determining test content and establishing content-related validity evidence are similar to those in the Standards (AERA et al., 2014). These standards are used as tools by the legal profession to assess to assess the validity of the test development and score interpretation process ("Gulino et al. v. Board of Education of the City School District of the City of New York," 2015; "Ricci et al v. DeStefano et al.," 2009; Sireci & Parker, 2006).

Table 2.1 provides a summary of relevant content from the Standards pertaining to this study (AERA et al., 2014). With regards to establishing test content, The Standards indicate the need to define test specifications (Standard 1.1, 4.1), including the purpose and target population as well as intended interpretation and use(s) of test scores (11.13). Of particular importance for credentialing tests, Standard 11.3 states:

When test content is a primary source of validity evidence in support of the interpretation for the use of a test for employment decisions or credentialing, a close link between test content and the job or professional/occupational requirements should be demonstrated (p. 178). In the absence of a good criterion and difficulties in collecting some other types of evidence, credentialing programs rely very heavily on content-related evidence. The Standards recognize this dependency, and provide guidelines to specify the need for clear documentation of the

rationale for method selection, procedures used, and SME qualifications and training (Standard 1.9, 1.11).

Table 2.1. Summary of Content-Related Standards for Educational and Psychological Testing

Standard #	Description
1.1/4.1	Requirements for test specifications (testing population, intended interpretations and uses)
1.9	Documentation of qualifications and training of subject matter experts and descriptions of the procedures in which they were involved
1.11	Documentation of procedures, including justification for use of importance indicators, used in determining test content in order to support alignment of test content with the sampled domain
11.2	Description of test content domain and alignment with content
11.3	Alignment of test content to specified domain when content is primary validity evidence
11.13	Clear definition of and rationale for content domain in support of test score interpretation and use when test purpose is credentialing

Similar to the Standards (AERA et al., 2014), NCCA accreditation guidelines (National Commission for Certifying Agencies, 2016) are not intended to dictate specific methods, but to provide guidance in selecting methods that will achieve a quality testing system and support fairness, validity, and reliability of test scores. NCCA standards, summarized in Table 2.2, require specification of the test purpose (1), documentation of the qualifications of panel experts as well as the decisions made by experts (13), identification and analysis of domains and tasks related to the professional role (14), and creation of a test weighting plan (blueprint) based on the practice analysis study (15).

Table 2.2. Summary of Content-Related NCCA Standards

Standard #	Description
1	Specify test purpose.
13	Panels must be composed of qualified experts in the given profession. Standard specifies the need for documentation of qualifications, responsibilities, and panel decisions/outcomes.
14	Job analysis study must be conducted and published, including identification/analysis of domains and tasks relevant to the role.
15	Establish exam specifications, including section weights that are established using a job analysis study method.

The second set of standards provided by an accreditation body is the ANSI 17024 (International Standard Organization Committee on Conformity Assessment, 2012). Requirements for documented test specifications and a documented job or practice analysis study are similar to those of NCCA. Standard 8.2 in Table 2.3 specifies the necessity for defining program scope, job, task, competencies, and abilities, eligibility requirements, and any applicable “code of conduct” (2012, p. 9). Standard 8.4e specifically calls for a practice analysis that specifies tasks and competencies necessary to perform defined tasks.

Table 2.3. Summary of Content-Related ANSI Standards

Standard #	Description
8.2	Certification scheme includes certification scope, prerequisites, job, task, competency, and ability (if applicable) descriptions.
8.4e	Required documents include job or practice analysis study.

The final set of rules guiding the establishment of test content and validity evidence are regulations residing in Title VII of the Civil Rights Act ("Civil Rights Act of 1964," 1964) that are governed by the Equal Employment Opportunity Commission (EEOC). The sections listed in

Table 2.4 describe the necessity of properly defining test content such that it aligns with specific job tasks in order to establish that a test used for employment purposes (i.e., hiring, firing, raises, promotions) does not discriminate based on race, color, religion, sex, or national origin. These regulations contain similar requirements stated in other standards, but emphasize the goal of avoiding bias in certain employment-related processes such as “hiring, promotion, demotion, membership (for example, in a labor organization), referral, retention, and licensing and certification, to the extent that licensing and certification may be covered by Federal equal employment opportunity law” (Equal Employment Opportunity Commission, August 25, 1978, Section 2B).

Table 2.4. Summary of Content-Related EEOC Statutes

Section	Description
1607.4D	Adverse impact as defined by the four fifths rule
1607.5	Content related evidence should demonstrate relationship between test content and job performance
1607.14B	Review of the job needed that includes a job analysis study
1607.14B2	Job review identifying important behaviors/tasks, avoiding possible bias.
1607.14B4	Samples used should be representative considering race, ethnicity, and sex, of the profession
1607.14C2	Requirements of job analysis in defining important behaviors and related tasks and relative importance.
1607.15C3	Documentation of job analysis procedures and results

According to EEOC guidelines, “any method of job analysis may be used if it provides the information required for the validation strategy used” (Equal Employment Opportunity Commission, August 25, 1978, Section 1607.14). Thus, if the goal is to provide content-related

validity evidence that links the content of the test with the specific tasks performed in a particular profession or job role (Sections 1607.5, 1607.14C2), then the selected method should support that goal and do so in a manner that avoids bias for select groups. This is accomplished by conducting and documenting a valid practice analysis that includes the use of representative subject matter experts from all subgroups in the target population (Sections 1607.14B, B2, B4, C2, and C3). This will be discussed in more depth in the next section related to legal defensibility.

Legal Defensibility

As Sireci and Parker (2006) specified in their review of validity evidence as a legal defense, there are our four legal bases for complaints against a testing program. First, Title VI and Title VII of the 1964 Civil Rights Act, equal protection under the Fourteenth Amendment (Sireci & Parker, 2006). Titles VI and VII address issues of discrimination of minorities as described in the previous section, where Title VI specifically relates to federally funded programs. As described previously, the statutes enforced by the Equal Employment Opportunity Commission (August 25, 1978) specify test or selection-based discrimination in terms of disparate impact and the four-fifths rule (Section 1604.7D). This rule states that if the pass rate for the specified subgroup (i.e., black/African- American) is less than the pass rate for the majority group there is evidence of disparate impact.

When evidence of disparate impact has been presented, the regulation states that test developers and/or users must demonstrate a legitimate business need for the credential and that a similar measure of the same job role that is less biased is not available (Equal Employment Opportunity Commission, August 25, 1978). Evidence must be presented regarding the methods chosen for determining content of the test and how the results provide an adequate link between

test content and competent job performance (Section 1607.5). Evidence is also needed to demonstrate adequate representation of minority groups within the test development process and efforts made to avoid bias.

Several examples of legal suites based on Title VII ("Civil Rights Act of 1964," 1964) were identified. In the most recent, *Gulino v. Board of Education of City School District of NY* (2015), a class action law suit was made related to two tests required for licensure in the state: ALST and LAST, two version of a test of liberal arts knowledge. Plaintiffs in this case contended that the tests were designed to measure knowledge required of university liberal arts courses as opposed to knowledge and skills needed for competent job performance. As the court determined the evidence presented to be sufficient in establishing clear disparate impact, the "pass rates for African-American and Latino applicants were between 54% and 75% of the pass rate for Caucasians" (2015, p. 2), the burden of proof shifted to the test developers and users to demonstrate that the test content was job related and the credential a business necessity.

The ruling in this case was that the test developers did not establish specific tasks related to teaching liberal arts, but maintained that testing general knowledge was sufficient for the test's purpose. According to this court, "for purposes of a Title VII discrimination action challenging an employment examination, to perform a suitable job analysis, a test developer must: (1) identify the tasks involved in performing the job; (2) include a thorough survey of the relative importance of the various skills involved in the job in question; and (3) define the degree of competency required in regard to each skill" ("*Gulino et al. v. Board of Education of the City School District of the City of New York*," 2015, p. 2).

The court deemed the test development procedures insufficient; tasks were not identified and the survey was considered insufficient to represent the views of minority groups even though

the percent of minorities in the sample were equivalent to what was believed to be the proportion in the population. The overall sample size was considered too small.

An older example that provided legal precedence for the Gulino case was *Guardians v. Civil Service Commission of the City of NY* (1980), also based on Title VII. In this example, the job analysis methods were deemed acceptable for establishing sufficient validity. The job analysis for the Guardians Association of NYC Police Department test for selection purposes included interviews, observations, expert panel review, and large sample survey (n=5600) with adequate representation of minority groups throughout the processes. The Defendants' still lost the suit, but not due to an inadequate job analysis. It was due to a lack of evidence in support of rank ordering test candidates based on scores established using a criterion-referenced test.

A final example was the United States Supreme Court case where the Plaintiffs sued the city of New Haven, Connecticut under Title VII when the city failed to certify results of a fire fighter promotion exam due to concerns over the low pass rate of minority candidates ("*Ricci v. Destafano et al.*, 557," U.S. Supreme Court, 2009). The test results were thrown out in hopes of avoiding a suit filed by minority fire fighters based on disparate impact after results indicated all African American fire fighters failed the test. In a 5-4 decision in favor of the Plaintiffs, the Supreme Court determined that the practice analysis method used to establish test content was based on necessary job tasks and was conducted using oversampling of minority groups in all test development processes to safeguard against bias. In short, the test developers adequately fulfilled the burden of proof to establish that the content was job related and another equally valid, less biased test was unavailable.

Beyond Titles VI and VII, two other legal bases for a test-related law suit exist (Sireci & Parker, 2006). First, the Fourteenth Amendment specifies equal protection under the law a for

all individuals (Sireci & Parker, 2006), which may be indicated in Title VI and VII cases where disparate impact exists for some minority group and they are denied access to a profession. Fourteenth Amendment is referenced in the *Guardian et al. case* ("*Guardians v. Civil Service Commission of the City of New York*," 1980), indicating that selection based on a list of rank ordered candidates derived from test scores "constituted discrimination against blacks and Hispanics in violation of the Fourteen Amendment, Title VII, and various other state laws" (p. 7). In this case, the issue was not test development process, but use of test scores.

The final legal basis for a suit against a testing program, due process under the Fourteenth Amendment, may not directly relate to the issues discussed regarding choice and application of practice analysis methodology. Due process relates to either procedural requirements of a testing program, such as providing advanced notice of testing requirements, or to substantive requirements. Buckendahl and Hunt (2005) described the courts application of substantive due process with relation to testing programs as "the relationship between the requirements of a test...and its purpose" (p. 150). In *Armstead v. Starkville Municipal Separate School District* (1971) a court of appeals ruled that a test of general knowledge, the GRE, administered to teacher applicants was not appropriate for use in employment selection, citing the Fourteenth amendment and stating that the test was "not reasonably related to the purpose for which it was designed" (Section 279, p. 4). No cases were found that directly link issues with due process to practice analysis methods.

The standards described herein and examples of potential legal issues whose outcomes were impacted by selected practice analysis methods are intended to provide some insight into why certain methods are chosen. In the following section, four methods commonly cited in the

literature are described as well as their strengths and weaknesses in fulfilling the requirements set forth in the standards and courts.

Review of Practice Analysis Methodology

There is more than one way in which to conduct a practice analysis study, as long as there is a defensible rationale for a given choice and the selected methods meet standards and legal requirements when applicable. Numerous methods have been developed over the past 60 years intended to guide experts in defining job roles and professions for the purpose of credentialing tests. Most of these methods originated from I/O Psychology procedures used to define job roles for the purpose of employee selection, placement, promotions, and raises (Raymond, 2001). These methods range from qualitative procedures that require collaboration between testing experts and SMEs to quantitative procedures using the results of a survey. Qualitative processes involve job analysis or testing experts to determine key tasks performed on the job from literature, observations, and/or interviews; or determine these tasks through collaboration with subject matter experts who are intimately familiar with a particular role.

Selecting an appropriate method for a given situation requires reflecting on the standards presented in Table 2.1 and the possible legal implications in the rare case that a testing professional is asked to defend the choice in court. As a primary source of validity evidence, a rationale for choice of methods should be established (AERA et al., 2014, Standard 11.13). The selected method should be feasible given available resources and time constraints.

Key texts on practice analysis methods describe critical incident technique (CIT) (Flanagan, 1954) and model-based approaches (Kane, 1997; LaDuca, 1994), but focus on the most common method of task inventory (Gael, 1983) and its variations (Clauser, Margolis, & Case, 2006; Knapp & Knapp, 1995; Raymond, 2001, 2005, 2016). Reviews of practice analysis

methods occasionally provide an example of a study, but rarely look to the applied practice analysis literature in professional journals to describe what is reflected in practice. Although there are limitations in generalizability among published studies as many studies conducted go unpublished, some insight into which methods are used in practice may be gleaned from examining these studies within the discussion of methodology. Five practice analysis methods (CIT, model-based approaches, Delphi method, SME conferences, and task inventory questionnaires) will be described as well as examples from practice analysis studies published in the last 10 years.

Studies provided as examples of current methodological selection and application originated primarily from peer-reviewed journals published within the past 10 years. These were found in database searches of Academic One and Google Scholar using keywords practice analysis, job analysis, job task analysis, role delineation study, critical incident technique, professional practice model, and task inventory questionnaire. The table in Appendix A. provides a list and description of studies that met the aforementioned criteria and were included in the following discussion.

CIT. One of the early methods for determining content of an employment-related test, the critical incidence technique (CIT), was established by Flanagan (1954) to define job roles in the field of I/O Psychology. Researchers often include this method as one approach to conducting a practice analysis for credentialing tests, but it is rarely seen in practice (Knapp & Knapp, 1995; Raymond, 2005). CIT is a way of articulating observed behavior into a description of a particular incident, specifically defining a successful or failed action. For example, an aviator may use a particular maneuver to avoid crashing a plane. That action would then be documented. Many hundreds or thousands of these incidences would be documented then

analyzed by a job analysis professional. CIT may be an appropriate method particularly in professions where performed tasks are often in response to context and client/patient-specific situations that may have clearly positive or negative outcomes. One example was found in the literature. Battenfield and Schehl (2013) conducted a practice analysis study for human performance technologist by conducting 44 individual interviews and five focus groups interviews driven by CIT-type questions as defined by Flanagan (1954).

The drawback to this method is the time and resources necessary to conduct such a study. This can be an undue burden for a credentialing program, particularly for small programs with limited resources (Raymond, 2005, 2016). In addition, it requires specific training for a testing professional to apply this method, both in collecting the incidents and then analyzing and transforming them into a content framework. Testing professionals outside of I/O Psychology are unlikely to have such training.

Model-based approaches. Similarly, model-based approaches are described in summaries of practice analysis methods (Knapp & Knapp, 1995; Raymond, 2005; Raymond & Neustal, 2006), but not commonly found in practice. The few examples found were not in journal-published studies, but on association websites in medical-related fields with more complex professional models that may not be supported by a task or procedure-driven model. LaDuca (1994) described an approach to defining a profession in terms of the professions social aspects or ecology, in addition to cognitive demands with the goal of creating a model of professional practice. Reflective of the CIT method where situations are analyzed as opposed to specific tasks, professional responsibilities are defined based on practitioners' self-reported patient encounters and other supporting documentation. A professional practice model is specified for complex professions where professional practice is highly situation-based such as

medicine and law (Hockberger, La Duca, Orr, Reinhart, & Sklar, 2003; 1994). Similar complexities to CIT exist, requiring large numbers of SMEs and long periods of time to collect and analyze data to sufficiently define and validate a complex model.

This specific approach was not cited in any of the studies listed in Appendix A. The most recent example found in published literature was a study conducted in 2003 for emergency medicine doctors (Hockberger et al., 2003). An initial model was developed using an advisory panel of clinical practitioners to determine specific clinical problems and encounters and task categories. In a second phase, the results of this panel were used to create a survey that was administered to a large sample of board-certified physicians followed by a SME panel using results to determine a final model of the profession.

Kane (1997) also described the model-based approach, similar to the professional practice model introduced by LaDuca (1994). In Kane's description, "patterns of practice" (p. 9) are defined by a group of SMEs that focus on client descriptions and needs, practice context and tasks, and the specific knowledge, skills, and judgments necessary to perform tasks in meeting client needs within the particular contexts. The outcome is a matrix rather than a list of tasks and KSAs that would be found in a task-driven approach such as the Delphi and task inventory methods. Kenward (2007) described a practice analysis study using Kane's model-based approach for establishing tasks and KSAs for nurse practitioners and clinical nurse specialists. An initial model was developed based on a review of the literature and then reviewed by a SME panel. The final model was validated using a survey.

Vandaveer, Lowman, Pearlman, and Brannick (2016) used a similar model-based approach to define the role of a psychology coach, which they referred to as a competency-model approach. An initial model was created based on a literature review and interviews with 27

SMEs, followed by a survey. The final model was determined based on the mixed methods results.

Model-based approaches have a number of challenges. First, they are time and resource intensive, requiring lengthy data collection procedures using large numbers of SMEs both remotely and in person. Second, the complexity of these models makes translating results into a test plan difficult and is not adequately described in the literature. If a survey is included, as in the previously cited study for nurse practitioners (Kenward, 2007), a blueprint may be generated, but the blueprint is likely to be very complex, thus difficult to construct a test that adequately aligns with the model. This leads to the final issue, which is that a more complex blueprint may require a more complex test (i.e., alternative item types, performance-based). As indicated in the literature, model-based approaches were designed for high-level medical licensure exams (Kane, 1997; LaDuca, 1994) where the added complexity is warranted. These methods, however, may not be reasonable for other industries; the lack of examples in published literature may support this conclusion.

Delphi. More common in examples of studies found in the literature are methods that focus on some process for determining essential tasks performed on the job and the KSAs necessary to competently perform those tasks. Although the Delphi method is not included in most summaries of practice analysis methodology (Clauser et al., 2006; Knapp & Knapp, 1995; Raymond, 2002, 2016; Raymond & Neustal, 2006), its function aligns with the goals of a practice analysis study using a repeated survey method to reach subject matter experts consensus on which tasks and/or KSAs should be tested (Hsu & Sandford, 2007). Surveys may be administered until consensus is reached, although research indicates three is typically adequate. Drawbacks to the technique are the length of time it may take given the need to iteratively

receive responses, evaluate responses, and make necessary changes. Participants may be slow to respond and some may drop out, reducing anticipated response rates. This could be particularly problematic for a testing program that is attempting to complete development on a reasonable timeline and with adequate representation of practice settings and critical subgroups.

Two related studies (Bevans et al., 2011; Castro et al., 2011) listed in Appendix A described using the Delphi method in a practice analysis for Clinical Research Nurses. They reported using the method to establish a content framework, beginning with an in-depth literature review to develop the content domains and tasks for the initial survey, followed by two additional surveys. The goal was to gain consensus on the essential tasks and overarching domains performed by research nurses. Once these were established, a large-scale survey was administered and reported in a separate study to validate the results.

SME conference. A SME conference requires a group of representative SMEs to determine the tasks and KSAs necessary for competent job performance (Gael, 1983, 1988a). Conferences were described as 3-4 day, in-person meetings with 6-7 SMEs and a facilitator collaborating to first develop a set of tasks and/or behaviors, as well as contingencies based on different job conditions, followed by knowledge and skills. Finally, a questionnaire was developed such that the participating SMEs would rate each of the tasks and KSAs. Although Gael's SME conferences are not specifically cited in the literature, the approach appears as a critical first step in most studies, particularly in conjunction with task inventory questionnaires. In fact, the majority of reported studies used some combination of an in person meeting of SMEs followed by a questionnaire, usually administered to a large sample of representative professionals. Knapp and Knapp (1995) described the function of a practice analysis advisory committee, using similar parameters to those described by Gael such as a SME panel of 8-10

representative participants who are responsible for defining tasks and KSAs in order to develop a questionnaire. Other researchers described similar SME groups using procedures for defining tasks and KSAs (Clauser et al., 2006; Raymond, 2002, 2016).

Current studies do not specifically state the use of a SME conference method, but most describe procedures similar to the method. The method is not used on its own in current studies, but typically in combination with other methods, typically survey methods. Weaknesses of this method on its own may be limited generalizability if the SME panel is not truly representative of the profession and limited utility of results given the few choices for statistical analysis with very small samples.

Task inventory questionnaire. A task inventory questionnaire method is the most commonly used method in practice (Raymond, 2002, 2016). The method was initially created to develop job descriptions within large corporations who needed to collect data from geographically dispersed employees (Gael, 1983). Procedurally, initial steps focus on development of a comprehensive list of tasks performed on the job. For credentialing tests, this is typically followed by the development of KSAs. In some cases, competency areas or job functions as opposed to tasks are defined (Battenfield & Schehl, 2013; McKenzie et al., 2016; Taub, Gilmore, Olsen, & Connell, 2015; Yang et al., 2012). It follows the same concept and procedures; simply a difference in how the job tasks are defined and labeled.

Following development of tasks and/or KSAs, a questionnaire is constructed that will ask practitioners to evaluate importance of each element based on one or more rating scales. This format allows the responding SMEs to indicate a level of agreement or disagreement based on a prompt such as, "How important is each task for an entry-level Montessori teacher?" Another

common question asked is “How often does an entry-level teacher perform the task?” In published studies, the survey portion is typically the focus.

In early studies tasks were created by job analysts who would glean information from previous job descriptions, followed by interviews with and observations of professionals (Gael, 1983). At end of his description of job task inventories, Gael described what he termed as a shortcut to this procedure: the SME conference, described above. Modern application of the task inventory combines this shortcut that utilizes experts in the profession to determine tasks and then verification of their work using a task inventory questionnaire. For example, Althouse, Du, and Ham (2009) conducted a two-phase study, the first of which was three-day meeting with 13 SMEs tasked with determining the competencies of a pediatrician. SMEs then indicated the importance and frequency with which a pediatrician performs each area. This process was followed by a survey of 1000 certified pediatricians asked to rate each of the competencies using the same scales.

Variations on a task inventory may incorporate additional research methods found in CIT and model-based approaches to support the work of the SMEs in defining tasks and KSAs. Similar to the work Gael (1983) described, methods include literature reviews, interviews, job logs, and observations. Many published studies indicated that a literature review was conducted prior to the work performed by SMEs, primarily used to provide an initial framework on which to base the task list before meeting with SMEs (Bradley et al., 2010; Castro et al., 2011; Erickson, Erickson, Campbell, Brekke, & Sandor, 2013; Gorman et al., 2010; Johanson, Miller, Coe, & Campo, 2016; Matarese et al., 2012; Starling et al., 2014; Tahan, Watson, & Sminkey, 2015; Vandaveer et al., 2016; Villanueva, Thompson, Macpherson, Meunier, & Hilton, 2006).

A number of studies used the results of previous practice analysis studies as a starting point for developing current tasks and KSAs (Althouse et al., 2009; Arbet, Lathrop, & Hooker, 2009; Garbin & Chmielewski, 2013; Gerrow, Murphy, & Boyd, 2006; Muckle, Plaus, Henderson, & Waters, 2012; Reuter-Rice, Madden, Gutknecht, & Foerster, 2016; Universal Public Purchasing Certification Council Governance Board, 2009; Willens, DePascale, & Penny, 2010; Zrebiec, 2014). A few studies also included interviews and job logs (Vandaveer et al., 2016; Zrebiec, 2014). For example, Willens et al. (2010) asked a panel of SMEs to review and revise an existing set of tasks prior to administering a survey. Gerrow et al. (2006); Taub, Olsen, Gilmore, and Connell (2008); and Tsai and Kramer (2014) conducted similar studies without using an initial SME panel. Tasks and KSAs from previous studies were used to construct the new questionnaire. In other instances, the process was reversed whereby a survey was administered using a questionnaire from a previous study, followed by a SME group review of the results in order to make final determinations about the content framework (Babcock et al., 2013).

Although results of the survey are most often the focus of reported studies, the development of tasks and KSAs are essential to the overall method. Selection of SMEs to participate in the process must ensure representation of the various aspects of a profession such as demographics and relevant practice settings (Knapp & Knapp, 1995; Raymond, 2002, 2016). Facilitation of the process is critical to ensure that the outcome represents a consensus among experts. Tasks and KSAs are typically developed using established formats that facilitate a clear understanding of the work processes and necessary knowledge, skills, and abilities to competently perform. The following descriptions address the issues of developing tasks and KSAs and developing a quality questionnaire.

Development of tasks, knowledge, skills, and abilities. Tasks define what a professional needs to be able to do to perform competently in a job role. Once tasks are established it is then necessary to determine what KSAs are necessary to perform these tasks competently. An example of this is presented in Table 1.1 above. Test content is not typically written based on the tasks themselves, but on the KSAs necessary to competently perform tasks. KSAs may be developed during an initial practice analysis study or after the study has been completed, depending on time and resources available, and whether KSAs will be included on the questionnaire.

KSAs are sometimes included in the questionnaire, but there is some disagreement in the literature regarding the appropriateness of this practice (LaDuca, 2006; Raymond, 2001; Wang, Schnipke, & Witt, 2005; Wang, Witt, & Schnipke, 2006). Two key issues are at play. First is the perceived cognitive complexity involved in indicating the importance of a particular piece of knowledge or skill for competently performing a task, compared with indicating whether a task is important for a job role. This is particularly problematic on a questionnaire where a respondent is asked to first rate a task followed by rating a KSA, requiring a shift in thinking in order to accurately interpret the response format.

The second and more prevalent issue revealed in the literature is the extended length of the survey questionnaire when KSAs are included (Raymond, 2001). Inclusion of KSAs may drastically increase the length of the questionnaire and reduce response rates. For those who do respond, responses may be less thoughtfully considered in the interest of completing the questionnaire. Some research has indicated that on longer surveys, participants tend to select the same response for all items near the end of a long survey as they become fatigued (Wang et al., 2005).

Constructing a questionnaire. In addition to general best practices for questionnaire design, there is a body of literature specific to practice analysis survey methods. Of particular interest in designing a task inventory questionnaire is the evaluation of rating scales and analysis of responses.

Selecting response formats or rating scales requires consideration of the profession and available information relevant in selecting the content for a credentialing test. For example, a test for licensure, the purpose of which is protecting the public, would require knowing which tasks and/or KSAs are most relevant to avoiding harm or reducing risk of adverse consequences for consumers (Raymond, 2005, 2016). Alternatively, content on a low stakes, voluntary certification test could be closely aligned with tasks that are most commonly performed. Other rating scales measure whether tasks are necessary for entry-level professionals, level of education required to be able to perform tasks, or general importance (Raymond, 2002, 2005). For example Bradley et al. (2010) and Keller et al. (2016) asked survey participants to indicate time to acquisition; meaning how long after starting practice did they take to acquire the specified competencies.

The most commonly used rating scales in published studies were importance and frequency, sometimes with an addition scale included (Althouse et al., 2009; Barnsteiner & Wyatt, 2002; Bevans et al., 2011; Erickson et al., 2013; Gorman et al., 2010; Johanson et al., 2016; Koby & Melby, 2013; Muckle, Apatov, & Plaus, 2009; Muenzen, Corrigan, Smith, & Rodrigue, 2005; Reuter-Rice et al., 2016; Tahan, Watson, & Sminkey, 2016; Villanueva et al., 2006; Willens et al., 2010; Wolever, Jordan, Lawson, & Moore, 2016; Yang et al., 2012; Zrebiec, 2014). Although these are often used together, experts and supporting research caution against using these two in combination as they are too highly correlated (Raymond, 2016).

Raymond emphasized the multidimensional nature of importance ratings, suggesting that people interpret the term to mean a combination of significance, frequency, and/or difficulty. Using importance alone may be acceptable, but in conjunction with other rating scales is problematic.

Research indicates the use of two to three rating scales is optimal, but more becomes cumbersome for the respondents and do not increase the precision of the final ratings (Raymond, 2005). This was reflected in the reviewed studies, with the number of response scales ranging from 1-3.

Analysis and creating a blueprint. Initial analysis of results may be as simple as reporting means and standard deviations to complex scaling of responses using a Rasch Rating Scale Model (RRSM) (Wang & Stahl, 2012) or Facets (Wang & Stahl, 2012). The ultimate goal of analysis is to create a test blueprint, thus, the steps to achieve this must be adequate to support the resulting weight given to the tasks and/or KSAs. There are three primary steps: 1) describe the data, generally reported as mean or median ratings and standard deviations or ranges, 2) combine rating scales if more than one is used, and 3) establishing the final weights for the blueprint (Raymond, 2005).

A fundamental issue with data generated from scales used in rating tasks and KSAs is that data are ordinal as opposed to interval, but are often treated as interval in the analysis (Harwell & Gatti, 2001; Spray & Huang, 2000; Wang & Stahl, 2012). Most published studies do not deal with this issue, nor do general descriptions of practice analysis methodology. However, in order to support the use of the blueprint as critical evidence of validity, this issue should not be ignored.

The primacy given to the opinions from the sample of professionals who respond to the questionnaires is clear in the published studies listed in Appendix A. In most cases, construction

of tasks and KSAs is only mentioned or given a brief description with little to no information regarding the experts or procedures involved. Studies present detailed accounts of the survey, subsequent analyses and conclusions. This is a heavy burden to place on individuals who may spend a limited amount of time responding to a questionnaire and may also have a limited understanding of what is being asked of them and the consequences of their responses.

Questioning current methods. The credentialing industry has exhibited strong faith in the adequacy of the task inventory questionnaire for the purpose of defining test content and providing essential content-related evidence of validity, as suggested in the previously cited literature. This includes both articles and texts written on practice analysis methodology as well as published practice analysis studies. Although an efficient method for accomplishing its goals that includes an empirical study that instills confidence in its practitioners, it is not without limitations. Colton, Kane, Kingsbury, and Estes (1991) specified four general assumptions of the task inventory questionnaire method: 1) participants are representative and qualified, 2) data analysis is done correctly, 3) the task list adequately describes the practice or job role, and 4) participants accurately interpreted and responded to the questions being asked. The accuracy of conclusions drawn based on survey results are, thus, based on the fulfillment of these assumptions. Colton et al. (1991) pointed to the specific issues with representation given typically low response rates on practice analysis surveys as well as the fourth assumption described above regarding accuracy of responses. Analysis of subgroup responses such as examining correlation of ratings among practitioners who work within a common setting may provide one method for examining accuracy, but it requires adequate sample sizes among subgroups in order to support inferences.

Babcock and Yoes (2013) also pointed to difficulties in determining the accuracy of a practice analysis survey. They used data from Centers for Medicaid and Medicare Services (CMS) to compare the job of a radiologist to information about services provided to patients. The analysis used a hierarchical linear modeling technique to compare survey responses to frequency of occurrences of particular tasks in the CMS over an eight year period. Although providing some evidence in support of survey results, it proved limiting in that data in the CMS may have reflected the job of an experience radiographer rather than entry-level. This procedure assumes large sample sizes and an external source of data such as a CMS database available to the medical field.

There have been two published studies that address the value added to evidence in support of validity by conducting a survey following the SME panel's work: Tannenbaum and Wesley (1993) and Maurer and Tross (2000). Tannenbaum and Wesley (1993) tested the research question "are the job analysis results from a carefully selected advisory committee of content experts the same as those obtained from a survey of the field" (p. 976), indicating that no published research had investigated the necessity of surveys in creating the content domain of a credentialing test. The first study (1993) used a teacher licensure test of foreign language with a 10-person panel developing 166 KSAs to be rated using a single importance scale. Three measures of agreement were used to compare the 7 survey responses from the panel and 423 survey responses: Pearson Product Moment Correlation, Intraclass Correlation (ICC), and a dichotomous measure comparing number of KSAs dropped from the domain based on a preset cut point. The results indicated a strong relationship between the two groups ($r = .85$, $ICC = .65$, dichotomous = 96% agreement).

The second study was also a teacher licensure test for chemistry, with an 8-member panel and survey sample of 800. The survey included a single scale of importance used to rate 181 knowledge statements. Similar to the first study, comparisons between the 7 panel members who responded were compared to 329 teachers in the larger sample suggested strong agreement ($r = .82$, $ICC = .62$, dichotomous = 96% agreement). Researchers concluded that the survey may not be necessary when a carefully selected panel is used to define the content domain.

Maurer and Tross (2000) conducted two similar studies, one that compared ratings on tasks and a second comparing rating on KSAs. These researchers suggest that the purpose of the survey is to provide information that you would not otherwise get from using a panel alone to determine the content domain of a test. This study was intended to demonstrate the similarity in ratings between the panel and larger group in order to provide evidence that the survey is not providing additional information. The test used for this study was an examination for personnel selection (candidates are within a single organization) as opposed to a licensure such as the one used in the previous example. Group comparisons were made using a combination of two scales, importance and frequency with Pearson's Correlation, ICC, a dichotomous index similar to the Tannenbaum and Wesley (1993) study, and finally Cohen's Kappa applied to the dichotomous judgments.

Sample sizes for this study were very small, with 9 panel members and 39 survey participants rating the tasks and another 8 panel members and 32 survey participants rating KSAs. Results of task comparisons were similar to the Tannenbaum and Wesley (1993) study, producing values that support strong agreement between the two groups ($r = .87$, $ICC = .83$, dichotomous = 85% agreement, and $K = .63$). Similar results were found when rating the KSAs with the exception of Cohen's Kappa ($r = .86$, $.84$, dichotomous = 97.3% agreement, $K = .38$).

Neither of these studies indicated potential flaws in the interpretability of their results. In both cases they report parametric statistics (means and standard deviations) and applied parametric statistical tests (correlation and ICC) that assume normally distributed data using data generated from ordinal rating scales. In addition to these data being ordinal, with the very small panel sizes and the small survey sample sizes in the Maurer and Tross (2000) study, data are unlikely to meet assumptions.

In addition, there was little to no description of the panel and survey demographics in these studies, making it difficult to assess how representative the panels were of the general populations. In addition, generalizability of the Maurer and Tross (2000) results is limited given that the job analysis was based on a personnel selection test within a single organization where the job role is likely very homogenous.

The current study to define the role of an early childhood Montessori teacher (ECMT) is intended to address some of these weaknesses and provide additional evidence in support of conducting practice analysis studies without a survey under conditions where the analysis is for a new certification program with a small population of professionals who work across diverse practice settings. Under such conditions, representation is more complex and careful panel selection is critical.

Similar to the Tannenbaum and Wesley (1993) the selected professional credential for this study is a specialty teacher certification. As described in the introduction, the Montessori teaching profession currently has a credential that is earned by attending a teacher-training program. However, there is no independent certification test similar to the Praxis[®] tests required for licensure in many states. The following description provides some key characteristics of

Montessori education and the current landscape, as well as benefits of a Montessori specialty certification for teachers, followed by the methods used to conduct the practice analysis.

Montessori Education

One might expect articles or book chapters on Montessori to begin with definitions of Montessori education. That is rarely the case. No single definition of Montessori education or even consensus on how to describe it in current day practice exists. Is it a method, a theory, a philosophy, pedagogy, and/or a system of education? For Maria Montessori, her work was not intended to be any of these. As a physician working with special needs children in the “slums of Rome,” her goal was to determine a way to educate these children (Kramer, 1976; Polk Lillard, 1996). The success of the method led to a “scientific system of education” (Montessori, 1948, p. 19) that has spread around the globe.

An overview of Montessori reveals a method that focuses on the whole child, encompassing education for the child’s intellectual, physical, emotional, and psychological development (Rambush, 1962). There are essential components of Montessori education as found in the writings of Maria Montessori. These include the fundamental ideas that there is an inextricable link between the development of physical movement and cognition, that it is important for children to have choices, and that intrinsic motivation and self-direction are critical to learning (Lillard, 2005). Materials in Montessori classrooms typically contain items specially designed by Maria Montessori to support these fundamental concepts. Montessori teachers’ fundamental task is to prepare a developmentally appropriate environment, which provides children with the structured freedom and opportunity for independent exploration and self-learning (Montessori, 1948). Focus on constant observation and assessment of students by teachers informs individualized instruction.

Another fundamental component of a Montessori classroom is mixed-age groups that are based on Maria Montessori's planes of development (Lillard, 2005; Polk Lillard, 1996). There are four planes of development: 1) ages 0-6, 2) ages 6-12, 3) ages 12-18, and 4) ages 18-24. Within each plane, there are two groups that define the typical classroom setup. For example, ages 0-6 are divided into 0-3 and 3-6, the latter being the range of ages for an early childhood Montessori classroom. Ages 6-9 constitute lower elementary, and 9-12 upper elementary. The designated age ranges are essential to the application of a core principle of the Montessori education: peer interactions in the classroom uniquely promote learning and development.

An example of Montessori teaching may vary a little or a lot depending on the school or classroom. It is applied in many more settings now than when it originally came to the United States in 1911 (Whitescarver & Cossentino, 2008) where it began as an innovation for wealthy children in private schools. Over 100 years later, Montessori has spread into public education, through charter, magnet, and mainstream public school classrooms (National Center for Montessori in the Public Sector, 2014).

The Montessori movement and key associations. As a formal movement, Montessori began with a single organization, Association Montessori Internationale (AMI), with Maria Montessori as its leader. Later her son Mario became its champion, with its key mission to protect the integrity of Montessori's work (Whitescarver & Cossentino, 2008). Later, the American Montessori Society (AMS) was created as the American arm of the AMI. The two groups parted ways in 1963 due to differences of opinion on how to adequately prepare teachers so that the community could meet the needs of an American system of education. While the two associations have a long history of tensions that make collaboration a challenge, Whitescarver and Cossentino (2008) in their description of today's Montessori movement point out that "both

organizations [AMS and AMI] view the extension of Montessori education to larger numbers of children as a key, mission-based priority” (p. 2591). In order to accomplish this mission, a common set of standards for schools and teachers would be beneficial.

The Montessori teaching profession is a difficult landscape to accurately describe. There is no overarching governance of the Montessori educational system and teacher-training institutions are not held to a single standard, although the latter is changing. Many Montessori teacher-training programs are now accredited by the Montessori Accreditation Council for Teacher Education (MACTE). Unlike most organizations in the disparate landscape of the Montessori community, MACTE has established a cohesive set of standards that has gained national acceptance. Its mission is to accredit teacher-training programs, both in the US and abroad, and is recognized as the accrediting body for American training programs by the US Department of Education (MACTE, 2015a). There are an estimated 184 MACTE accredited teacher training programs in the US (MACTE, 2015b) with additional programs in the process of becoming accredited. Most of these programs are aligned with a Montessori affiliate, an organization that has its own teacher-training program and certificate based on a unique set of standards that meet the MACTE requirements. Table 2.5 presents a list of the seven affiliates that maintain MACTE accreditation for some or all of their programs.

Table 2.5. MACTE Accreditation Affiliates

Association	Number MACTE Teacher Training Programs
American Montessori Society (AMS)	93
American Montessori International/USA (AMI)	14
International Montessori Council (IMC)	6
Montessori Education Programs International (MEPI)	7
Pan American Montessori Society (PAMS)	3
International Association of Progressive Montessori (IAPM)	2

The current state of training and teacher certification as well as the relationship to a hypothetical industry certification is presented in Figure 2.1. The path to certification is presented at the top, indicating that this path what MACTE accredits. Each of the seven affiliates listed in table 2.5 issues a separate, independent credential indicating that candidates have met the individual program requirements. An industry certification, as suggested in this study would be independent of each existing credential, establishing a unified standard.



Figure 2.1. Summary of path to certification for Montessori teachers.

MACTE standards for teacher accreditation. Similar to the Council for Accreditation of Education Preparation (CAEP), which accredits traditional teacher-training programs, MACTE has established minimum requirements that Montessori teacher training programs must meet in order to achieve accreditation (MACTE, 2015a). For example, early childhood teachers from a MACTE accredited program must complete a minimum of 200 hours of coursework and 400 hours of supervised practice teaching in order to complete a program and attain a teaching credential (MACTE, 2015b). A set of established teacher competencies are indicated in the accreditation standards that define the content areas covered by coursework and assessments.

The application process requires a training program description presented to MACTE in the form of a self-assessment that incorporates an overall summary of the program and descriptions of evidence in support of three principles of quality: 1) candidate learning, 2) faculty learning and inquiry, and 3) program capacity (MACTE, 2013). In addition to self-assessment, a MACTE representative conducts an on-site visit to verify the program has met all requirements.

One of the challenges that MACTE and the larger Montessori community face is the lack of recognition by states of the Montessori teacher credential (MACTE, 2015b). In most states, the Montessori-specific training is not recognized as fulfilling the educational requirements for licensure, and there is no Montessori-specific, industry-level certification, such as the one presented in Figure 2.1. For teachers who hold a Montessori credential from a Montessori training program, this does not preclude them from teaching in a private or public charter school in states where licensure is not required of teachers in these schools. However, it does preclude them from teaching Montessori in traditional public schools, magnet schools, and charter schools where state licensure is required. It also limits the professional status of the Montessori teacher in the eyes of the public and larger teaching profession.

Advocacy efforts by MACTE and state Montessori organizations are beginning to make some progress in gaining acceptance of Montessori-trained teachers. South Carolina established an alternative path to licensure for Montessori teachers who have earned a bachelor's degree or higher, and who have graduated from a MACTE accredited program. In addition, South Carolina requires Montessori teachers to pass the appropriate Praxis® tests (content and pedagogy) based on the age for which the teacher is credentialed. The testing portion of the credential is problematic for a couple of reasons. First, the age ranges tested in the different versions of the Praxis® Principles of Teaching and Learning do not align with those of the Montessori credential and the age groupings in Montessori schools (Educational Testing Services, 2015). Second, the pedagogy tested on the Praxis® does not align with Montessori pedagogy. Although passing the test meets the requirements of state licensure, it does not reflect competency in a Montessori classroom.

As the Montessori community works toward establishing recognition of its teacher workforce, development of a certification test that mirrors the existing pedagogical tests for traditional teachers may be beneficial in addressing these issues. In addition to aiding advocacy efforts, a voluntary certification for Montessori teachers working in both private and public classrooms, may enhance the quality of Montessori instruction and give teachers a tool for promoting themselves as professionals. This test could also serve as an add-on test for attainment of state licensure in states where Montessori advocates have gained acceptance of the credential.

Summary of benefits of a Montessori-specific certification test. There are many possible benefits to developing a certification test for the Montessori community of teachers, students, and parents, as well as for states. A Montessori certification could initially be voluntary

with the possibility of becoming a requirement of licensure in the long run. The following benefits include those for all stakeholder groups:

- Provides an objective measure of competency, independent of specific training, for those hiring Montessori teachers.
- Provides an additional verification of competency for teachers to use in marketing their skills.
- Provides a potential path for traditional teachers who want to transition to Montessori and need a way to gain recognition.
- Provides a possible alternative to the Praxis[®] Principals of Learning and Teaching (PPLT) or other teacher pedagogy test (i.e., Child Development Associate credential for early childhood teachers) that align with traditional classroom pedagogy and grade levels.
- Provides a mechanism for professionalizing Montessori teachers and encourages more teachers to enter the field.
- Supports consistent application of Montessori teaching across different types of classrooms (i.e., public charter, public mainstream, private).

Applying practice analysis methods to defining the role of a Montessori teacher. For the current study, the first step to achieving these benefits and establishing a Montessori teaching certification was to select one particular role and conduct a practice analysis. This would provide the necessary project on which to evaluate the research questions for this study. As indicated in the introduction, the role of an early child Montessori teacher was selected for this study. Because the concept of an industry certification is novel to the Montessori community and resources were limited for the project, the scope was limited to ECMTs in the state of South Carolina. The following chapter describes the specific method used to conduct this study.

CHAPTER 3. METHODS

A professional practice analysis, using a combination of a subject matter expert (SME) conference and task inventory method described in the literature in Chapter 2, was conducted in order to address the stated research questions. This included a three-phase process: 1) planning, 2) a one-day, in person meeting, and 3) a survey. Phase I, planning, was used to recruit SMEs for two groups, an advisory group (AG) and a larger panel of SMEs for the in-person meeting. Also during Phase I, the advisory group assisted with creating initial materials for Phase II. Phase II of the study focused on an in-person meeting with 13 SMEs tasked with creating the content framework (domains and tasks) that would then be used to develop a questionnaire for the survey in Phase III. This final phase was the development and administration of a questionnaire that would provide the data to answer the study's research questions.

The following section describes the methods in more detail, beginning with a general description of participants and how they were recruited. Next, procedures for each of the three phases will be described.

Participants

Participants of this study were early childhood Montessori teachers (ECMTs), ECMT trainers, and administrators who hire and oversee ECMTs. Volunteers were recruited through several avenues: public and private schools, teacher education programs, and the South Carolina Montessori Alliance. At the beginning of the project, one volunteer was recruited to act as a Montessori advisor and liaison with the Montessori community. The advisor aided in making connections with Montessori organizations (e.g., SCMA, teacher education programs, private

schools), assisted with the advisory group, and helped to organize and facilitate the panel meeting in Phase II. This advisor, a credentialed ECMT, Montessori teacher trainer, Montessori administrator, and policy professional was recruited in North Carolina but had connections with the Montessori community in South Carolina.

Identifying potential participants. A database of schools (see Appendix B) was created from two sources: 1) list of South Carolina schools provided by The Riley Institute (2014) that have ECMTs, and 2) the SCMA (2015) website, which lists all known Montessori schools in South Carolina. These two sources were combined and revised over the course of the study, with schools being added or removed depending on whether the school was still open, operating at least one early childhood classroom with a credentialed teacher, and reachable. In addition, snowball sampling was used, particularly for teacher education programs communicating information regarding study participation to graduates and teacher trainers.

The final database of schools included 24 public schools in 15 districts and 21 private schools. Table 3.1 contains a summary of these schools and districts as well as the estimated number of teachers and the percent of public and private school ECMTs included in the study. The first column, number of schools in SC with early childhood programs, includes all schools in the list prior to contacting any public districts or schools followed by the numbers actually participating in the study. At this point, participation of a public school in the study simply meant that invitations to participate in one of the three phases of the study could be distributed to school administrators and/ or teachers per district approval. Private schools were contacted at this point only if there was no current email address or website. In this initial step of creating a database, there was no commitment from districts, schools, or administrators that their teachers would actively participate in the study.

Table 3.1. Summary of Sampling by Schools and ECMTS

	Schools with EC programs		Schools included in the study		ECMTs in schools		ECMTs in the study	
	<i>N</i>	%	<i>n</i>	%	<i>N</i>	%	<i>n</i>	%
Private	25	41.7	21	46.7	102	44.3	90	48.6
Public	35	58.3	24	53.3	128	55.7	95	51.4
Total	60	100.0	45	100.0	230	100.0	185	100.0

Eighty-four percent of private schools in the original list were included in the initial database (21 of 25 identified schools) and 68.6% (24 of 35 identified schools) of public schools. Public schools made up 53.3% and private schools 46.7% of the study sample. Similarly, public ECMTs constituted 51.4% and private ECMTs 48.6% of the teachers in the sample.

Private programs and ECMTs not included in the database were from schools where contact could not be made. To include public schools in the database, district offices were first contacted in order to attain permission to communicate directly with schools. Three districts had formal processes for applying to do research: Richland One (Columbia), Richland Two (Columbia), and Charleston. Richland One and Charleston districts denied the request to do research in their districts due to the number of studies already in progress in their districts; thus, schools in these districts were removed from the study (total of 28 teachers). These districts were limiting academic research studies in the fall of 2015 due to large numbers of missed school days attributable to severe weather. Due to the recruitment efforts of SCMA and teacher training programs, one teacher from Richland One participated in the panel meeting, which was conducted on a Saturday (outside of school time and location), and three teachers from Richland One responded to the survey.

In addition to Richland One and Charleston districts, Anderson 2 and 4, and Berkley districts were not included (number of schools = 3, ECMTs=5). Berkley district office indicated that their one Montessori classroom was being eliminated and permission could not be attained for Anderson districts.

The last four columns in Table 3.1 provide the number of estimated ECMTs working in public and private Montessori programs. Ninety, or 88.2% of the estimated number of private school ECMTs in SC, were included in the study database. Ninety-five or 74.2% of the estimated number of public school ECMTs in SC were included.

Recruitment of participants. Two primary documents (see Appendix C) were distributed via email to districts, schools, and organizations to inform administrators and teachers about the study and to solicit volunteers for all three phases. SME qualifications were initially assessed by phone and email prior to their participation in the study and then final demographics collected in the questionnaire administered in Phase III, so that identical descriptive information would be available for both the panel and the broader population.

In order to participate in Phases I and II, SMEs had to meet the following requirements:

- Hold an ECMT credential from a MACTE accredited teacher preparation program
- 3-8 years as an ECMT or ECMT educator
- Currently working as an ECMT or ECMT educator

The years of experience requested was somewhat arbitrary, intending to recruit experts who would have enough experience to understand the role of an ECMT, but not so far from an entry-level teacher that understanding the role from that perspective would be a challenge. Some compromise was made, particularly with regards to experience, to ensure representation of the four known affiliates in South Carolina (AMS, AMI, MEPI, and IND). One panel member was a

first year teacher, representing AMI trained teachers. Teacher trainers tended to have more years of teaching experience, but were justified for inclusion given that they work on a daily basis preparing teachers for this role. Detailed demographics of the panel are provided in the results chapter.

Recruitment of participants for Phases I and II were completed in October and November of 2015, focusing on the districts and schools with the largest number of ECMTs and targeting the three geographical (upstate, midland, and lowlands) and population (urban, suburban, and rural) regions of the state. Table 3.2 indicates the target plan for representation of the population with regards to practice settings and affiliations. Practice settings describe the different environments in which teachers work that may impact the types of tasks they do or implementation of those tasks (i.e., prepare a classroom, assess student learning). Lines 5 and 6 in the table indicate Montessori classroom or classrooms within a traditional public school. A public non-charter (line 7) is a public school that is all or mostly Montessori. For example, the Lexington 4 Early Childhood Center is an entirely Montessori public school. Latta Elementary is also a Montessori public school, but has one optional traditional classroom at each grade level. Both schools fell into this category. There is currently only one Montessori charter school (line 8) in the state. One teacher was recruited from this school, but was unable to attend the panel meeting.

Table 3.2. Recruitment Plan for Panel SMEs

	Practice Setting	Teacher/School Affiliation
1	Private	AMS
2	Private	AMI
3	Private	IND
4	Private	MEPI
5	Public school within school	Any
6	Public school within school	Any
7	Public non-charter	Any
8	Public charter	Any
9	Teacher trainer	AMI
10	Teacher trainer	AMS
11	Teacher trainer	IGS-MEPI
12	Teacher trainer	IND

An individual’s affiliation may be defined in terms of the credential they hold, the training program that employs them, or the school in which they teach. For the purposes of this study, affiliation was defined by the credential they held, meaning where they received their training. If an individual received training at Lander University, for example, they would have an AMS affiliation. For teacher trainers the affiliation of the program in which they are training teachers was considered in addition to where they received their training. Therefore, a teacher may hold an IND credential, but teach in an AMS program such as Lander University.

In addition to geographic regions, affiliation, and practice settings, participants of a practice analysis study should be representative of other demographic characteristics in the population such as race, ethnicity, and sex. Little is known about the demographic of Montessori teachers in general and no data exists for these teachers in the state of South Carolina. However, effort was made to recruit as diverse a panel as possible.

Participants of the AG were recruited first so that work could begin immediately on establishing information and materials to be used by the panel. Five SMEs participated in the AG: two public ECMTs, three private school ECMTs, and one administrator. One of the public

teachers was also employed as a teacher trainer. MEPI, AMI, AMS, and IND affiliates were represented in the advisory group. Each participant was asked to sign a consent form for participating in the study (Appendix D).

Phase I

The AG group was initially recruited with the intention of having three one hour, web-based meetings. Only two were held, on November 10, 2015 and December 8, 2015, with all necessary work completed in those two meetings. Agenda and minutes from these meetings are included in Appendix E. From these discussions, initial information needed for the panel meeting were generated and reviewed. These included a list of possible purposes of the hypothetical certification test; the description of a just qualified ECMT; proposed eligibility requirements; five proposed content domains; and sample task, knowledge, and skill statements.

Phase II

Fourteen SMEs were recruited and 13 attended the one-day panel meeting (one SME dropped out due to illness). Each participant was required to sign a consent form for participation in the study (Appendix F). The panel consisted of nine full-time ECMTs, three full-time teacher trainers from three different affiliated training programs (AMS, IND, and MEPI), and one administrator. All panel SMEs were credentialed Montessori teachers or administrators. In addition to the three full-time teacher trainers, five ECMTs on the panel also work as part time teacher trainers. ECMTs represented four affiliates (AMS, AMI, IND, and MEPI) and had an average of 7.17 years of experience ($SD = 3.639$, range = 1-12).

Panel SMEs were primarily female ($n = 11$, 84.6%). Two of the panel members were non-white: one Hispanic and one Asian. Almost all held at least a bachelor's degree (11 of 13)

with eight having a master's degree and one a doctoral degree. Additional demographic information is presented in the results chapter.

Panel meeting. The meeting was held on Saturday, January 10, 2016 at the Montessori School of Camden in Camden, SC. During Phase I, content for a workbook (Appendix G) was developed for use in the panel meeting that included the agenda, glossary of terminology, documents developed by the advisory group, and instructions for writing task statements. Workbooks were sent to the panelists one week prior to the meeting and also provided in print on the day of the meeting.

In order to fully understand the target population for the test, the AG had developed a list of eligibility requirements, minimum qualifications that must be fulfilled or actions taken prior to sitting for a test, which included certification from a MACTE accredited program, minimum 400 hours of practicum work, and recommendation of a bachelor's degree. Although the panel concurred that the proposed eligibility requirements could be used for guiding this research study, a future certification is hypothetical and such requirements would need to be addressed by a credentialing association or governing board. They noted issues with limiting the definition to only those who hold a credential from a MACTE accredited program as well as the lack of requirement for a bachelor's degree. The panel indicated that a credential that could be accepted by the state as part of licensure would need to require a bachelor's degree.

Development of tasks. SMEs initially worked in four small groups, predetermined based on practice setting, experience, and affiliation. Within small groups, they drafted and categorized informal task statements based on the five domains proposed by the AG. An additional category was provided for this exercise for tasks that did not align with the existing domains. The initial informal tasks are listed in Appendix H.

The small groups were provided one and one-half hours to create these initial tasks. After completion, the entire group reviewed and revised the tasks into formal statements. SMEs were instructed to focus on the essential tasks specific to the work of a just qualified, entry-level ECMT. The panel discussed each of the tasks from the perspective of their diverse practice settings and experience to come to consensus on which tasks would be included on the final list. Eighty-one initial tasks were reduced to the final 19 presented in Table 4.1, in the Results chapter. Each task statement was revised to follow a standard task format (Raymond, 2005) with a verb, direct object, materials used to accomplish the task, and purpose of the task. This format would support the work to be accomplished in the next step in the process of developing a content framework, writing KSAs. That exercise was outside of scope of this study, but was considered when developing the tasks so that a future program may use the results of this study.

The final formal tasks were reviewed and consensus reached on all 19. These tasks provided the content framework needed to develop a questionnaire for the Phase III survey. Next, the panel reviewed and approved the wording of the proposed rating scales to be used on the questionnaire described in Phase III.

Finally, participants completed a brief questionnaire intended to evaluate the quality of the meeting and its outcomes. Questions were adapted from an existing questionnaire (Johnson, 2008). Results from 11 of the 13 participants who chose to respond to the questionnaire are summarized in Table 3.3. The full feedback form can be reviewed in Appendix I.

Table 3.3. Summary of Practice Analysis Meeting Feedback

Question	Median (<i>n</i> = 11)	Minimum (<i>n</i> = 11)	Maximum (<i>n</i> = 11)
Orientation was clear	4	3	4
Leader clearly explained tasks	4	4	4
Training helped understand tasks	4	3	4
Group discussion aided understanding	4	4	4
Equal opportunity to contribute	4	4	4
Utility of pre-meeting materials	4	2	4
Confident of defensibility of work	4	3	4

Responses indicate that the panel considered the purpose and tasks of the meeting clear, that training was sufficient to aid their understanding of the process, group discussions were useful and there was equal opportunity for everyone to contribute ideas and opinions. A couple of people did not find the pre-meeting materials helpful, but the consensus was that the overall process was defensible. Participants were also given the option to make comments or feedback for specific changes. The few comments that were made were complementary of the work and facilitation, with three participants suggesting a longer meeting with more breaks. Overall, the panel successfully accomplished the goal without complications or concerns expressed.

Phase III

A validation census survey method was used to establish comparison data between the panel of experts and the general population. The sample described earlier included all identified ECMTs in South Carolina whose school was willing to receive invitations to the study. In addition, ECMT trainers and administrators in the state were invited to participate. The following description includes the development, pilot testing, revision, and final administration of the questionnaire.

Development of the questionnaire. The online pilot questionnaire was developed using Qualtrics software (Qualtrics, 2009). Questionnaire content (Appendix J) included instructions, a summary of the target audience and eligibility requirements, a task rating section divided by the five domains and two scales, and demographic questions. Two rating scales, Criticality and Frequency, were selected (Raymond, 2005, 2016). The wording of the rating scales was reviewed at the end of the panel meeting, allowing the SMEs to select the format that would be most easily understood by the target population. They were presented with two scales each for measuring criticality and frequency. These are listed in Table 3.4, with the chosen options highlighted. SMEs agreed that in measuring criticality, the word “harm” should not be used. Public education, in particular, has an established definition focused on physical harm to students, making the more diverse meaning used in the task ratings potentially difficult to comprehend. Thus, the scale that referred to adverse consequences was chosen to measure criticality. SMEs also reviewed two options for measuring frequency, selecting option 1 in Table 3.4.

Table 3.4. Rating Scale Options Presented to SME Panel

<p>Criticality option 1: Is there risk of harm if an entry-level ECMT does not perform the task competently? Harm may be physical, mental, emotional, or financial.</p>					
No risk of harm					Very high risk of harm
0	1	2	3	4	5
<p>Criticality option 2: Is there risk of adverse consequences if an entry-level ECMT does not perform the task competently? Adverse consequences include physical, mental, emotional, or financial.</p>					
No risk of adverse consequences					Very high risk of adverse consequences
0	1	2	3	4	5
<p>Frequency option 1: How frequently does an entry-level ECMT perform the task?</p>					
Never perform					Perform very often
0	1	2	3	4	5
<p>Frequency option 2: How frequently does an entry-level ECMT perform the task?</p>					
Not responsible for					Several times per day
0	1	2	3	4	5

Demographic questions were selected based on suggestions in the literature; examples from published practice analysis studies, as well as factors related specifically ECMTs in South Carolina such as geography and practice settings. Twenty distinct questions were asked, 10 of which were general demographic questions. The remaining 10 were specific to job role (ECMT, teacher trainer, and administrator), thus, not all respondents would be asked all questions. Except for sex, ethnicity, race, age, and school responses to demographic questions and ratings were required by the software to avoid missing data.

Pilot test. Teachers, teacher trainers, and administrators from one public and one private school were invited to participate in pilot testing the questionnaire ($n = 30$). Invitations were sent by email (Appendix K). Selection was based on size of the school and willingness to participate in the pilot. The public program selected had the largest number of ECMTs in the state, as well

as teacher trainers and one administrator willing to participate. Participants of the pilot study were given one week to respond.

The pilot version of the questionnaire included a section at the end with four statements about the instrument’s clarity and ease of use (scale of 1-6) as well as an open text field for providing specific responses. There were 16 responses, slightly over a 50% response rate. Fourteen responses were from public and two from the private school. Fourteen respondents were ECMTs and 2 administrators. Median ratings and minimum and maximum values are reported in Table 3.5. Although the median ratings were high for all four statements, closer examination of the responses in conjunction with comments made at the end of the pilot questionnaire indicated the need to increase clarity of the instructions, particularly with the criticality options. Several of the respondents appear to have been confused about whether 0 was the high or low end of the scale. This issue was addressed in the revision for the final survey.

Table 3.5. Summary of Pilot Evaluation Questions

Statement	Median	Min	Max
The purpose of the study was clearly stated.	6.00	3	6
The instructions for how to complete the questionnaire were clear.	5.00	1	6
I understood how to rate each task using the criticality and frequency response formats.	6.00	4	6
The online questionnaire software (Qualtrics) was easy to use.	5.00	1	6

Communication and administration of final questionnaire. Revisions were made based on the pilot administration feedback (see final survey in Appendix L). Instructions for the criticality rating exercise were revised and a specific example provided to improve clarity. In addition, the formatting of the rating section of the questionnaire was revised. In the pilot, a side

by side format was used that allowed respondents to rate each task on the two scales at once. The rationale for this design was to reduce the amount of reading time, thus, the overall time needed to respond to the questionnaire. In this design, the ratings were displayed in a dropdown menu, rather than radio buttons placed horizontally across the page. In order to make the ordered responses more obvious, the format was changed to a horizontal display with radio buttons. In the revised format, all tasks were rated for criticality first, then frequency.

Once revisions and pilot testing of the final questionnaire were complete, a separate version of the same survey was created and sent to the AG and panel SMEs. This version required the participants to provide a first and last name, unlike the general survey, to ensure that only the appropriate people would be included in the panel ratings for analysis. These SMEs were sent an email invitation on Wednesday, January 27, 2016 and given five days to respond.

The final questionnaire was launched on February 1, 2016. In addition to earlier communications that were sent during the initial participant recruitment for the first two phases, an email announcement was sent to school administrators one week prior to the questionnaire launch. The list of schools and districts, along with number of ECMTs and city are presented in Appendix B. The survey announcement provided general information regarding the benefits of an ECMT certification to various stakeholder groups, information about the study, and instructions for participation. A follow up email was sent the day prior to the survey launch with the link to the questionnaire and access instructions. Both communications are included in Appendix M.

In addition to the emails sent by the researcher, a representative of the SCMA sent an invitation to their member list and a representative of Montessori Now, an advocacy group, posted the announcement and link to the questionnaire on its website and Facebook pages. To

control for non-representative respondents, two screening questions were asked at the beginning of the questionnaire:

- 1) Do you live in South Carolina?
- 2) Do you work in early childhood Montessori education (teacher, teacher educator, or administrator)?

Method for creating a test blueprint. In order to answer the questions for this study, the data first needed to be analyzed and transformed into a content specification, or blueprint, as described in the literature review. Every task on the list generated by the SME panel was included on the questionnaire after unanimous approval by the panel; thus, general importance was already determined. The next step was to determine the relative importance of each task (Lunz, Stahl, & James, 1989). The end result was a series of weightings reported as percentages of test content that would be represented on the test. It is assumed that tasks with higher ratings on criticality and frequency should have primacy on a test of ECMT knowledge and skills compared with tasks that have lower ratings. The first step, then, was to determine how to combine the two rating scales in a way that reflected the purpose of the test (Raymond, 2002, 2016; Spray & Huang, 2000).

There are many ways to do this. An additive model is a standard method where the responses to the two scales are simply added together: criticality + frequency (Raymond, 2002). A second option is a multiplicative model (Kane, 1997) that requires a judgment regarding which scale is more important and then multiplying that by some denomination or raised to a determined power (i.e., $\text{Criticality} \times \text{Frequency}$, $\text{Criticality}^2(\text{Frequency})$).

Weaknesses exist for both the additive or multiplicative models. Mathematically, additive models violate rules of combining different variables such as X and Y. Multiplicative models fail

to address difference in variances and covariances between the different scales (Raymond, 2016). Multiplying scale values may mask important differences within one scale that is indicative of different perspectives across demographics or practice settings and should be reflected in blueprint weights. This issue may be exacerbated when multiplying one scale by some factor or exponent.

Appendix N contains tables comparing b parameters, standard errors, and percent weights for four different models. In addition, reliability indicators (Appendix O) were reviewed as well as the scale structure, which will be presented in the results. Percent weights were very similar across the different models, with the ranking option being the most distinct. Reliability indexes were very similar and all within acceptable ranges. The deciding factor was the issue that three of the four options altered the ordinal nature of the scale, particularly when criticality was increased by multiplying by a factor or exponent. Thus, a multiplicative option was selected using the following formula to combine ratings from the two scales:

$$\sqrt{(\text{Criticality} + 1) * (\text{Frequency} + 1)}$$

The square root allowed the data to maintain the original six-point ordinal scale structure and one was added to criticality to ensure that a zero rating on criticality did not force a zero weight for the task.

One of the issues brought to light in both the practice analysis and survey methods literature is the limited application of data from ordinal scales produced by a Likert-type rating scale (Harwell & Gatti, 2001; Lunz et al., 1989; 2002; Spray & Huang, 2000; Wang & Stahl, 2012). Many of the statistical analyses applied to the data, including simply reporting means and standard deviations, imply an interval scale. In this study, the choice was made to only anchor the end points of the rating scale in order to achieve results that more closely mirror an interval

scale (Cook, Heath, Thompson, & Thompson, 2001; Tony & Klockars, 1982). However, in order to conduct the parametric analyses, particularly correlation analysis, to compare the panel ratings with the survey ratings, a Rasch Rating Scale model (RMSM) was used to place all the items on a single interval scale with a mean of 0 and standard deviation of 1, and then a linear transformation applied to attain the weights for each task. The model for RRSM is (Andrich, 1988):

$$\ln(P_{nik}/P_{ni}(K-1)) = B_n - D_i - C_{ik}$$

where,

P_{nik} = probability of rater n rating task i with category k

P_{ni} = probability of rater n rating task i with category (k-1)

B_n = theta for rater n

D_i = task importance measure for the i^{th} task

C_{ik} = task category measure to obtain k^{th} rating for i^{th} task

This formula takes the natural log of the probability that a rater selects one category (k) or another (k-1) for all possible categories. Theta is the item response theory (IRT) term for ability or, in this case, attitude or opinion. D is the measure that, in this study, will be referred to as the b parameter, the standard IRT term for item difficulty. This term is interpreted in the current context as the point on the theta scale where a person of that ability or strength of opinion has a 0.5 probability of selecting a particular category (C) on the rating scale. Since there are multiple categories, there are multiple points on the scale, or b parameter. Those reported in forthcoming results tables are the mean b parameter across all the categories. This will be presented in more concrete terms in the results.

Once parameters were estimated, a linear transformation was performed to convert the b parameter for each task to a percent weight in order to define the blueprint. The following transformation equation was suggested by Spray and Huang (2000).

$$[(-b) + |(\min(-b))| / \text{Sum}((-b) + |(\min(-b))|)] * 100$$

The reverse of the difficulty parameter was added to the absolute value of the minimum b parameter across all tasks, creating a task weight. All weights were summed and then each weight divided by this sum and multiplied times 100 to create a percent weight. These weightings are reported in the following results chapter.

CHAPTER 4. RESULTS

To provide a clear picture of how the research questions are answered, the results are reported within the framework of the three phases of the study. Answers to the research questions are a culmination of those three phases. Phase I presents the definitions of the entry-level ECMTS, test purpose, and five content domains developed by the AG. Phase II provides the content framework that was developed by the panel. Finally, Phase III survey results are presented and comparisons made between the Panel and Survey groups in order to answer the first two research sub-questions. Recall the content of these questions:

1. What are the final task ratings and blueprint weights based on responses to a questionnaire by the selected panel of subject matter experts?
2. How do the final task ratings and blueprint weights generated from surveying representative professionals compare to those of the panel?

The final blueprint is presented and comparisons for the two groups are presented using nonparametric analysis with the unscaled ratings followed by parametric analysis with the scaled ratings and blueprint weights.

Results of Phases I and II

The results of Phase I are the definitions of an entry level early childhood Montessori teacher (ECMT), the proposed test purposes, and the initial five content domains that guided the development of the task statements in Phase II. The first to be developed was the initial definition of a just qualified, entry level ECMT, which was generated by AG experts in the first web-based meeting. This definition follows.

An ECMT is an entry-level teacher who has graduated from a MACTE accredited early childhood preparation program, which requires a minimum 200 hours of academic work and 400 hours of practicum work. Teachers are prepared at a minimum in areas of philosophy, theory, and child development, practical life, sensorial, mathematics, language, science, physical geography, cultural studies, physical education, music, and arts education.

The ECMT has experience as a supervised teacher during the practicum portion of their education but is not expected to have unsupervised experience in the ECM classroom. A just qualified ECMT will have knowledge of the basic materials in a Montessori classroom.

This statement was intended to be a draft. It was reviewed in Phase II by the SMEs attending the in-person meeting.

The AG proposed three purposes for an ECMT certification program:

- A voluntary credential for all teachers that may be used to demonstrate competency and differentiate among them professionally, as well as for hiring managers to use as a verification of competency independent of the teacher preparation programs and affiliate-specific credentials.
- A mandatory certification used by state agencies to assess competency of an entry-level ECMT seeking licensure to work in ECM public education programs.
- Provide cohort-level feedback for teacher preparation programs in evaluating the competency of their graduates.

During the in-person meeting, the Panel agreed with all three purposes and added “demonstrating competency for families” and “increasing professionalism for ECMTs” to the list. Although these were noted, they are both potential benefits of the first two purposes as opposed to purposes of the test.

The five content domains were initially identified through group discussion during the first AG web-based meeting. These were then refined, with descriptions drafted by one AG volunteer. These were distributed prior to the second web-based meeting for review and comments, then discussed and finalized during that second meeting.

The final 19 tasks, developed by the in-person panel, were gathered into the five domains of practice to reflect broader, more manageable concepts for users of the test outline (for example, sub-scores. Final Domains and Tasks are presented in Table 4.1. A minor adjustment was made to one Domain title by the Panel. Domain 3 was initially Behavior and Classroom Management. Panel SMEs decided behavior management is part of classroom management, thus, these were combined.

Table 4.1. Content Domains and Tasks

Domain 1	Instruction: Providing instruction that is developmentally appropriate for each student in each curriculum area based on assessment of individual learning needs.
Task 1.1	Present individual and small group lessons across all curriculum areas using Montessori materials and activities to promote acquisition of skills and/or concepts.
Task 1.2	Adapt lesson presentation and environment based on individual needs, including learning differences, as determined by observation in order to promote development.
Domain 2	Prepared environment: The classroom environment that is developmentally appropriate for the physical, cognitive, and social-emotional development of the multi-aged group of students.
Task 2.1	Establish procedures using tools such as a daily schedule, grace and courtesy lessons, expected procedures, ground rules, and demonstration of lesson to support independence in the child.
Task 2.2	Sequence materials in each curriculum area using established Montessori guidelines to facilitate methodical acquisition of concepts and skills.
Task 2.3	Create original activities using Montessori guidelines in order to meet the needs of children with specific learning styles and/or interests.
Task 2.4	Maintain materials and the environment by repairing and restoring to ensure that all works are complete and ready for use by the children.
Task 2.5	Prepare the environment using Montessori guidelines to create appropriate curriculum areas, work spaces and activities in order to foster movement, order and independence.
Task 2.6	Continually assess the functionality of the classroom environment using observation in order to meet the changing needs of the children.
Domain 3	Classroom Management: The use of management techniques, the

	Montessori Peace curriculum, lessons of Grace and Courtesy, and the preparation of the environment to provide guidance and support for the social-emotional development of the mixed age group resulting in a respectful classroom community with positive social interactions.
Task 3.1	Guide a child to purposeful work through invitation to a lesson or activity in order to encourage engagement and normalization.
Task 3.2	Redirect a child who is engaged in disruptive, destructive, or dangerous behavior in order to encourage engagement and normalization.
Task 3.3	Establish ground rules and limits in cooperation with the children, using lessons such as grace and courtesy and community building, in order to encourage engagement and normalization.
Task 3.4	Manage transition times using auditory and visual signals in order to encourage independence and normalization.
Task 3.5	Facilitate large group activities (i.e. games, storytelling, walking the line, silence game, music enrichment) in order to establish a classroom community.
Task 3.6	Create a classroom where children can resolve conflicts, using methods such as a peace curriculum, for the purpose of developing classroom community.
Task 3.7	Assess the children’s engagement level in the environment using observation in order to determine whether to begin interaction and/or lesson demonstration.
Domain 4	Assessment: Evaluating students using formative (i.e., observation, check lists) and summative assessments in order to teach to individual learning needs.
Task 4.1	Assess the child using observation (i.e., running records, anecdotal, time sampling, check lists) to evaluate interest, progress, and subsequent presentations.
Task 4.2	Maintain a system of record keeping and planning to document the progress of each child’s presentation, practice, and mastery of skills and concepts.
Domain 5	Communication and Interaction: Engagement in interactions with children, parents, co-workers, and professionals, using a variety of means and methods, in order to communicate, educate and collaborate in a positive manner that supports the developmental needs of each student in the multi-age group.
Task 5.1	Communicate with parents on a regular basis through a variety of means (i.e., parent/teacher conferences, parent education, newsletters, reports, email, phone) in order to encourage a seamless transition for the child between home and school.
Task 5.2	Establish a partnership with assistant(s) and/or co-teacher(s) using communication tools (i.e., team meetings, education, job descriptions, set expectations) in order to create a consistent and productive learning environment.

There was a high level of consensus among the panel SMEs. All tasks aligned with the domains recommended by the AG, and no issues were noted where SMEs could not agree on essential tasks across the practice settings and affiliations. A “parking lot” was created at the

beginning of the process for tasks that did not fit into the proposed model and/or tasks that could not be agreed on by all SMEs. This category was never used.

The one area that was briefly touched on that may have been an area of derision between public and private Montessori classrooms was in relation to standardized assessments that are administered to young children. The panel agreed that these assessments are outside of the specific role of an ECMT and should not be included in a test of Montessori pedagogy. The issue would likely be more problematic in future practice analysis studies for lower and upper elementary teachers as standardized testing is a more significant educational component that impacts teaching and learning.

Phase III: Survey Results

As described in the methods, two versions of the survey questionnaire were developed and administered. These were identical except for the initial questions that asked those responding to the Panel version to include names. The Panel version of the questionnaire was administered to participants of both the AG and the Panel given that SMEs from both groups were involved in the planning and design of the content framework. All 16 SMEs from this group completed the questionnaire.

Eighty-six ECMTs, teacher trainers, and administrators responded to the questionnaire. Two cases were removed. One respondent indicated s/he was not a teacher, teacher trainer, or administrator, thus not part of the target audience for the survey. The second participant response removed was a teacher trainer from North Carolina, also not part of the target audience. This left 84 respondents for analysis. The two groups are referred to in the analysis as Panel and Survey response groups.

Response rate. Due to the lack of a database of direct contact information for the survey population, response rates are reported here as percent of schools responding and the percent of teachers based on the estimates gathered in the planning of this study. Table 4.2 presents the number of schools that had participating teachers compared to the estimated numbers of schools. Number of districts represented in the sample is also reported. Finally, number of administrators is also reported, with the estimated number of administrators in the sample equivalent to the number of schools. Schools were assumed to have one administrator of the early childhood Montessori program.

Table 4.2. Summary of Responses to the Survey

	Estimated Sample Size	Responses to Survey		Panel Total Count*	Total participation (Panel + Survey)	
		Count	%		Count	%
Private Schools	21	7	33.3	2	9	42.9
ECMTs	90	10	11.1	6	16	17.8
Admins	21	4	19.0	1	5	23.8
Public Schools	24	18	75	2	20	83.3
Districts	15	12	80	2	14	93.3
ECMTs	95	49	51.6	6	55	57.8
Admins	24	17	70.8	0	17	70.8

Note. Panel Total Count reflects the number of schools beyond those that participated in the survey so that schools are only counted once.

Districts reflect those who agreed to allow contact with their schools for participation.

Ten of the 90 estimated ECMTs (11.1%) responded from 7 of 21 private schools (33.3%). Forty-nine of the 95 estimated public ECMTs responded (51.6%) from 18 of 24 public schools (75%) invited to participate. Eighty percent of districts who agreed to participate in the study were represented. Recall from Table 3.1 that 46.7% of identified early childhood

Montessori programs in the state were private and 53.3% were public. Although there are slightly more ECMTs in public Montessori schools than private, these results are skewed toward public Montessori programs. In addition to schools and ECMTS, 4 private and 17 public administrators responded, also skewed toward public Montessori schools and programs.

The final three columns in Table 4.2 are intended to provide a high level view of project scope. Panel Total includes the number of responses from the panel and additional numbers of schools are those beyond the ones included in the survey. Total participation indicates that, after considering both Survey and Panel groups, 47.6% of private Montessori schools and 83.3% of public Montessori schools invited to participate in the study were represented. For ECMTs, 17.8% of private teachers and 53.7% of public teachers participated in the overall study.

Recall in the description of the survey questionnaire that the question asking for school was not a forced response. Sixteen of the 84 survey respondents did not provide this information. Schools were inferred from IP addresses for 3 of these respondents and 13 remain unknown. Of these 13, 1 was a private ECMT, 3 public ECMTs, and 9 administrators from public schools. In other words, the number of schools represented in the study is likely higher than what is reported here.

One final point regarding the responses summarized in Table 4.2 is that responses from Richland One schools are not included. They were not part of the original survey sample and these schools were never contacted directly to participate in the study. The demographic and analysis descriptions include these respondents, but they were not considered in estimating the response rates.

Initial screening of survey responses. Data were reviewed for missing values and outliers. Given the small sample size, screening was done primarily by visual investigation of the

data. Median, minimum, and maximum values were reviewed for each task. Ratings were summed across the 19 tasks for each participant in both groups to create a total score. Medians and minimum and maximum scores are presented in Table 4.3 for both criticality and frequency scales. There was one outlier in the panel data, a SME who had reversed the order of the Criticality scale. A follow up interview with the panel member confirmed that the individual had intended for the scores to be on the high end of the scale, thus, those responses were revised with the SMEs consent.

Table 4.3. Summary of Total Scores for Panel and Survey Groups

Total Score	Panel (n = 16)			Survey (n = 84)		
	Median	Min	Max	Median	Min	Max
Frequency	76.0	66	93	89.0	21	95
Criticality	85.0	60	95	80.0	0	95

In general, Criticality had a higher median score and broader range of scores than the Frequency scale. Panel scores were lower on Frequency and higher on Criticality. Total scores were also reviewed visually with boxplots (Figures 4.1 and 4.2). There were five respondents in the survey data that had extreme responses on one scale (total score on Criticality < 7). Removing responses from a practice analysis data set must be done with careful consideration, particularly when the sample size is very small as in this study. Removing individuals who represent various subgroups may negatively lessen adequate representation. However, given the evidence from the pilot study that the Criticality scale was confusing to some SMEs, it is likely that these five extreme cases experienced the same issues with the response format. In this case,

the responses would not accurately represent the opinions of the SMEs, providing justification for removal of these outliers.

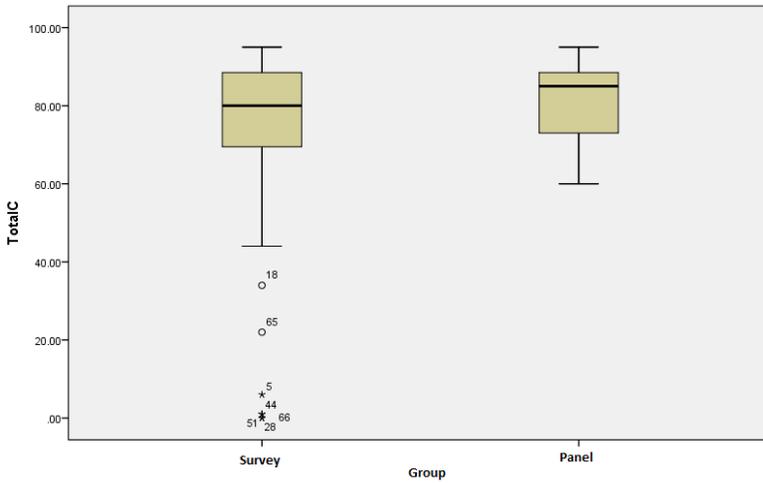


Figure 4.1. Distribution of criticality ratings for survey and panel scales.

Review of summed responses to the Frequency scale presented three extreme outliers (Figure 4.2). One was removed (66) due to the issue with Criticality ratings; however the two remaining extreme outliers on the frequency scale (27 and 45) were not removed. There was no similar justification for doing so.

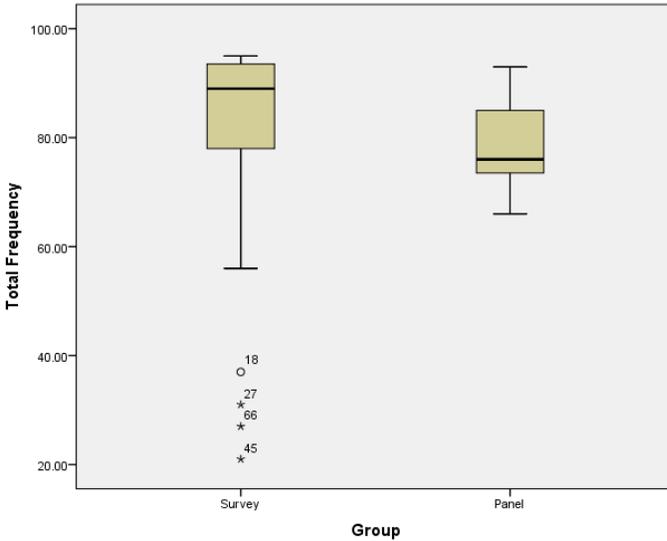


Figure 4.2. Distribution of frequency ratings for survey and panel scales.

Overall analysis was also done with removing various subgroups in order to determine the impact of removing them from the sample. Weights were estimated using procedures described in Chapter 3 and compared using all data, data with all extreme values removed, data with administrators removed, and data with non-certified survey participants removed. This analysis is summarized in Appendix P. These subgroups were singled out as they may represent subgroups of responses that are less representative of the target population. Weights were nearly identical across all groups, indicating that a blueprint developed from any of these sets of data would be the same, not impacting the outcome of this study. Thus, no additional cases were removed. The final sample size for analysis was 79.

Demographics. Demographic questions were included in exactly the same form on both the Panel and Survey questionnaires and results are presented here in side-by-side comparisons to aid in answer the overarching research question regarding whether the absence of a survey would negatively impact the strength of the validity argument. Representation of the profession is critical in order to infer an accurate blueprint from a practice analysis study. Quality of results

in either set of ratings is dependent, in large part, on representation of the population in both groups. To depend on the Panel results alone, evidence would need to be presented indicating adequate representation of demographics, experience, and practice settings.

Demographics, however, are self-reported and, thus, data presented here are the best indication available of representation in the profession. Some questions may have been misunderstood or incorrectly answered. Among panel responses, errors were discovered and corrected with participant permission related to questions of affiliation and status as teacher educators (trainers).

To begin the analysis, a comparison of the general demographics (sex, ethnicity, race, and age) between the two groups was conducted. Table 4.4 presents the frequencies and percentages of these four survey questions. As descriptions of the ECMT population are not available in the published literature, interpretation is limited to how the results compare between the two groups and the general population. The vast majority of SMEs in both groups were female (Panel = 87.5%, Survey = 97.6%) and white (Panel = 75%, Survey = 88.6%). Black participants made up 12.5% of the Panel and 5.1% of the Survey responses, and one individual in each group indicated Hispanic/Latino background.

Table 4.4. Comparison of General Demographics between Panel and Survey Respondents

Sex	Panel		Survey	
	Frequency	Percent	Frequency	Percent
Male	2	12.5	2	2.5
Female	14	87.5	77	97.6
<i>Total</i>	16	100.0	79	100.0
Ethnicity				
Hispanic/Latino	1	6.3	1	1.2
Not Hispanic/Latino	15	93.8	78	98.8
<i>Total</i>	16	100.0	79	100.0
Race				
Asian	0	0.0	1	1.3
Asian/White	1	6.3	0	0
American Indian/ Alaskan Native	0	0.0	1	1.3
Black	2	12.5	4	5.1
White	12	75.0	70	88.6
Other	1	6.3	3	3.8
<i>Total</i>	16	100.0	79	100.0
Age				
Under 25	0	0	2	2.5
26-35	6	37.5	14	17.7
36-45	3	18.8	28	35.4
46-55	3	18.8	20	25.3
Over 55	4	25.0	14	17.7
Missing	0	0.0	1	1.3
<i>Total</i>	16	100.0	79	100.0

The most common age range among Panel participants was 26-35 and for the Survey group 36-45. The Survey group had more participants in the mid ranges of 36-45 (Panel = 18.8%, Survey = 35.4%) and 46-55 (Panel = 18.8%, Survey = 25.3%). Age was included in this study, because it is a common demographic included in practice analysis studies; however, age is not necessarily indicative of experience and knowledge of the profession. Years of experience reported in the next few sections may provide better insight into representation of the profession.

Geography. Panel and Survey participants were compared across geographical and population regions of the state (Table 4.5). Information from the South Carolina government website (2016) was used to estimate numbers of schools and teachers in the geographical regions of the state: Upstate, Midlands, and Lowcountry. Representation of Upstate in both Panel and Survey groups were fairly close to the percent of teachers in that region (25.0% and 21.5% compared with 22.2% in the state) and very close in the Survey group in the Midlands (46.8% compared with 49.2%) and Lowcountry (31.6% compared with 28.7%) regions. However, Midlands represented a larger percentage (68.8%) of the total sample in the Panel group and much less in the Lowcountry (6.3%). This may be due to the location of the meeting that took place in the Midland region outside of Columbia, which was selected as a central location for participants coming from different parts of the state. In addition, the interest in this research was lower among private schools, and there is only public Montessori program in the Lowcountry that is not part of the Charleston School District, which denied the request to do research.

Table 4.5. Comparison of Geographic and Population Regions between Panel and Survey Respondents

Geographical Region	Panel		Survey		State	
	Frequency	Percent	Frequency	Percent	Frequency	Percent
Upstate	4	25	17	21.5	41	22.2
Midland	11	68.8	37	46.8	91	49.2
Lowcountry	1	6.3	25	31.6	41	28.7
Total	16	100	79	100	185	100
Population Region						
Urban	2	12.5	6	7.6	NA	NA
Suburban	7	43.8	24	30.4	NA	NA
Rural	7	43.8	49	62.0	NA	NA
Total	16	100	79	100	NA	NA

Regional comparisons are also presented graphically in Figure 4.3. Over-representation of the Midland region in the Panel group is clear in this image as is the lower representation in the Lowcountry. These distinctions may be less important, however, than representation of practice settings, which will be discussed later in the demographic analysis.

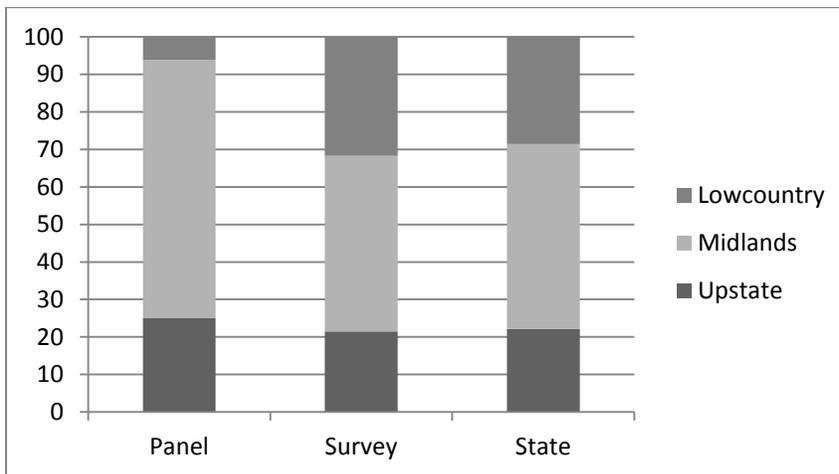


Figure 4.3. Comparison of panel, survey, state ECMTs by geographical region.

A similar comparison with state-level data was not feasible for population region. Although the United States Census Bureau provides a definition of urban and rural areas that could be used to categorize schools, a definition of suburban is more complicated. It appears from a review of the data that participant responses did not align with any formal definitions and, thus, comparisons to such data would be misleading. It is apparent that a large part of the state is rural (State of South Carolina Government, 2016) and major urban regions are Columbia, Charleston, Greenville, and Rock Hill (United States Census Bureau, 2016).

Strictly comparing the two groups, there were small numbers of responses from urban areas. Nearly half (43.8%) of panel participants reported being in suburban regions of the state compared to 30.4% from the Survey group, although according the Census Bureau's definitions, four of those seven participants from the Panel who indicated suburban actually live in rural areas. Sixty-two percent of teachers in the Survey group reported being from rural regions compared to around 43.8% of panel participants. Representation from urban regions was low in both groups with two (12.5%) representatives on the Panel and six (7.6%) in the Survey group.

Demographics related to experience, credentials, and practice settings were analyzed based on the specific job roles of ECMT, teacher trainer, and administrator. There is significant overlap among these groups, with over half of ECMTs indicating, that they were also working as teacher trainers and some also working at administrators for Montessori programs in their schools. Responses by ECMTs are reviewed next.

ECMTs. Table 4.6 presents summaries from both Panel and Survey ECMTs on questions related to education and years of experience. ECMTs who served in the Panel group had at least an Associate's degree, 25% a bachelor's degree, and half indicated a master's degree. A small number of Survey participants had less than an Associate's degree ($n = 2$, 3.4%), 37.3%

indicated having a bachelor’s degree, and 55.9% had a master’s degree. The one Panel participant indicating “Other” education had studied classical guitar, and the one Survey participant indicated “education specialist.”

Table 4.6. ECMT Comparison of Education and Experience

Education	Panel		Survey	
	Frequency	Percent	Frequency	Percent
Some college	0	0.0	2	3.4
Associate Degree	2	16.7	1	1.7
Bachelor's Degree	3	25.0	22	37.3
Master's Degree	6	50.0	33	55.9
Other	1	8.3	1	1.7
Total	12	100.0	59	100.0
Years of experience				
0-3 years	3	18.8	16	27.1
4-7 years	2	12.5	18	30.5
8-11 years	4	25.0	9	15.3
12-15 years	3	18.8	7	11.9
More than 15 years	0	0.0	9	15.3
Total	12	100.0	59	100.0

Panel participants had more years of experience with 25% indicating 8-11 years and 18.8% indicating 12-15 years. This is particularly relevant since the original recruitment of experts attempted to find ECMTs with approximately 3-8 years of experience, but finding teachers to attend a full day meeting was a challenge, thus more experienced teachers were recruited. This is not of particular concern for this study. The rationale for using the 3-8 year experience range is that as teachers become more experienced they may be less familiar with the tasks of an entry level professional. However, four of the seven Panel participants (57.1%) who

indicated more experience also indicated that they were ECMT trainers, suggesting familiarity with essential tasks necessary for entry-level teachers.

Over half of Survey responses indicated 0-3 or 4-7 years of experience. There were also a large number of very experienced teaches in the Survey sample, with nearly half indicating 8 or more years of experience. Similar to the Panel group, 14 of the 25 ECMTs (56.0%) who indicated more than 8 years of experience also indicated being teacher trainers. Table 4.7 presents a summary of additional roles that participating ECMTs fulfill. Fewer ECMTs fulfill the dual role of teacher and administrator, with one Panel participant (8.3%) and three Survey participants (3.4%) indicating that they also functioned as administrators of a Montessori school or program.

Table 4.7. Additional Roles of the ECMTs

	Panel		Survey	
Teacher Trainer	Frequency	Percent	Frequency	Percent
Yes	6	50.0	37	62.7
No	6	50.0	22	37.3
Total	12	100.0	59	100.0
Administrator				
Yes	1	8.3	2	3.4
No	11	91.7	57	96.6
Total	12	100.0	59	100.0

Montessori affiliation. Representation of affiliation is emphasized here as the current Montessori credentials are aligned with these separate affiliations. Establishing an industry standard would require consensus across these associations. As described earlier, affiliation may be aligned with the training program attended, training program in which the ECMT teaches, or school in which the ECMT is working. The affiliation requested from ECMTs was specified to

be affiliation related to the training program attended. Thus, 50% of Panel participants and 62.7% of Survey participants, as reported in Table 4.8, indicated AMS affiliation, specifying that they attended an AMS affiliated training program, earning an AMS credential.

Table 4.8. Summary of Education Affiliation of ECMTs

Education Affiliation	Panel		Survey	
	Frequency	Percent	Frequency	Percent
AMS	6	50.0	37	62.7
AMI	2	16.7	3	5.1
MEPI	3	25.0	9	15.3
IND	1	8.3	8	13.6
IMC/CGMS	0	0.0	1	1.7
None	0	0.0	1	1.7
Total	12	100.0	59	100.0

The bias toward AMS in both samples is justified due to its dominance in the United States as the largest Montessori training affiliate; it accounts for 63.8% of all MACTE accredited programs (MACTE, 2015a), and 40.0% of MACTE accredited training programs in South Carolina. Although there is no MACTE accredited AMI program in the state, there is one AMI training program in the state and a small representation from AMI trained teachers participated in both groups (Panel = 16.7%, Survey = 5.1%).

Affiliate representation is more clearly delineated in Figure 4.4, where Panel and Survey responses were visually compared. Individuals in the Panel group were carefully selected to represent AMS, MEPI, IND, and AMI programs. Although AMI is not an accredited program in the state, a truly representative sample for an industry certification would need to incorporate some representation of all affiliates and, thus, were included in this study.

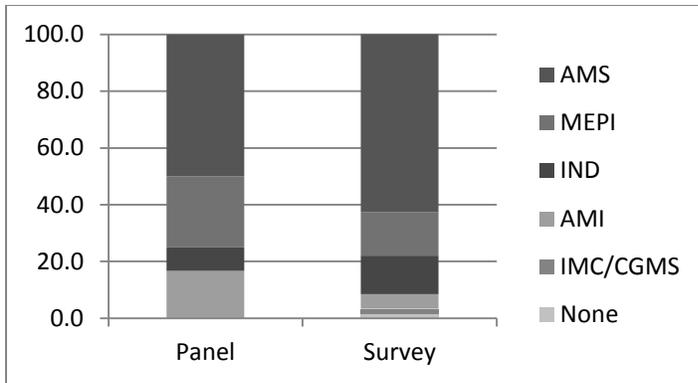


Figure 4.4. Comparison between panel and survey participant affiliation.

EMCT credentials and practice setting. Finally, credentials held and practice settings in which ECMTs work were compared, as shown in Table 4.9. All Panel participants were ECMT certified, one of whom held an additional credential as an Elementary 1 teacher. Approximately 92% of Survey responses came from credentialed ECMTs, with another three (5.1%) credentialed as an Elementary 1 teacher and two (3.4%) who indicated not having a Montessori credential. These last two participants were state licensed teachers working in an early childhood Montessori classroom. Fifty-eight percent of Panel participants and 84.7% of Survey participants were also state licensed. Given the high rate of response among public school ECMTs (see Table 4.2) and administrators a large number of state licensed teachers is to be expected.

Table 4.9. Comparison of Professional Certifications and Practice Settings

Montessori Certification	Panel		Survey	
	Frequency	Percent	Frequency	Percent
Early Childhood	11	91.7	53	89.8
Elementary 1	0	0.0	3	5.1
Early Childhood and Elementary 1	1	8.3	1	1.7
Not Certified	0	0.0	2	3.4
Total	12	100.0	59	100.0
State Licensed				
Yes	7	58.3	50	84.7
No	5	41.7	9	15.3
Total	12	100.0	59	100.0
Practice setting				
Private Montessori school	6	50.0	10	16.9
Public Montessori classroom within a traditional school	4	33.3	25	42.4
Public Montessori charter school	0	0.0	3	5.1
Public Montessori school (non- charter)	2	16.7	20	33.9
Public Montessori magnet program	0	0.0	1	1.7
Total	12	100.0	59	100.0

Similar to affiliation, representation of practice settings was essential to defining an industry credential that would be applicable in both public and private Montessori schools and programs. Survey responses failed to adequately represent the most essential categories, private and public. The Survey sample was heavily biased toward public ECMTs ($n = 49, 83.1\%$), with 10 responses (16.9%) from private ECMTs. A few responses from a Montessori Charter and a public Magnet were collected in the Survey group that were not available in the panel, but given that these do not represent the majority of Montessori programs in the state their lack of representation on the panel has very limited impact on the generalizability of the results to different practice settings. The Panel group included 50% public and 50% private Montessori

programs, with representation of both Public Montessori within a traditional classroom ($n = 4$, 33.3%) and Public Montessori schools ($n = 2$, 16.7%). There was no representation from the one Montessori Charter school in the state.

Teacher trainers. Table 4.10 presents a similar comparison to the previous section for those participants who indicated that they are ECMT trainers. The numbers of EMCT trainers is large. This may be due to the different training model that Montessori uses compared with traditional schools of education where many training programs use classroom teachers to teach courses (i.e., online, summers, weekends) and to supervise teachers-in-training who are fulfilling practicum hours. For example, Lander University has two full time teacher trainers on staff, and then uses classroom teachers to teach additional courses throughout the year (Irvin, 2015). It is also possible that some teachers in the Survey group misunderstood the question, creating a larger than expected number reported.

Table 4.10. Comparison of Teacher Trainers' Level of Education, Educational Affiliation, Years of Experience, and Credentials

	Panel		Survey	
	Frequency	Percent	Frequency	Percent
Education				
Some College	0	0.0	1	2.4
Associate Degree	0	0.0	1	2.4
Bachelor's Degree	0	0.0	18	43.9
Master's Degree	6	85.7	19	46.3
Doctoral Degree	0	0.0	1	2.4
Other	1	14.3	1	2.4
Total	7	100.0	41	100.0
Education Affiliation				
AMS	5	71.4	25	61.0
AMI	1	14.2	3	7.3
MEPI	1	14.2	5	12.2
IMC/CGMS	0	0.0	1	2.4
IND	0	0.0	6	14.6
None	0	0.0	1	2.4
Total	7	100.0	41	100.0
Years of experience training ECMTs				
0-3 years	0	0.0	14	34.1
4-7 years	2	28.6	14	34.1
8-11 years	1	14.2	7	17.1
12-15 years	2	28.6	3	7.3
More than 15 years	2	28.6	3	7.3
Total	7	100.0	41	100.0
Certification				
Early Childhood	6	91.7	36	87.8
Elementary 1	0	0.0	1	2.4
Early Childhood and Elementary 1	1	8.3	2	4.9
Not credentialed	0	0.0	2	4.9
Total	7	100.0	41	100.0

Six of the seven Panel teacher trainers indicated holding a master's degree, whereas Survey participants were divided between bachelor's (45.5%) and master's (45.5%) degrees, and

two participants indicated less education. Individuals in both groups who responded “Other” indicated training as Education Specialists.

The breakdown of education affiliation in Table 4.10 was very similar to the overall group, with the largest participation from AMS trained teachers, followed by MEPI and AMI. A larger percentage of Panel participants indicated more years of experience, all reporting 8 or more years. In comparison, more than half of ECMT trainers in the Survey group indicated 0-7 years of experience. All ECMT trainers in the Panel group and most of the Survey group ($n = 38, 92.7\%$) were credentialed ECMTs.

The question regarding affiliation in this section asked teacher trainers to indicate the affiliation of the training program in which they taught. These responses are summarized in Table 4.11, which includes an additional column for comparison with the proportion of training affiliates represented in the state. No AMI teacher trainers participated on the panel, but three responded in the larger survey. All three teacher trainer affiliates were represented on the panel (AMS, MEPI, and IND), with greater representation from AMS and IND programs. Although MEPI has two MACTE accredited training programs, making up 33.3% of the programs in the state, both are part of the same training organization (Institute for Guided Studies), so only one representative participated in the panel. The Survey group was more heavily representative of AMS, with much smaller representation from the other affiliates, including IMC/CMGS which does not have a training program based in the state.

Table 4.11. Affiliations Represented by Teacher Trainers

Teacher Education Affiliation	Panel		Survey		MACTE in SC	
	Frequency	Percent	Frequency	Percent	Frequency	Percent
AMS	3	40.0	27	65.9	2	33.3
AMI	0	0.0	3	7.3	0	0.0
MEPI	1	20.0	5	12.2	2	33.3
IMC/CMGS	0	0.0	1	2.4	0	0.0
IND	3	40.0	5	12.2	2	33.3
Total	7	100.0	41	100.0	5	0.0

Administrator demographics. The same analyses were conducted for administrators (see Table 4.12). Most administrators in both Panel and Survey groups had a high level of education, with all participants having at least a bachelor’s degree. Two of the three administrators on the Panel were also ECMTs and had a master’s degree. The additional administrator held a Doctorate. Half of the Survey participants who indicated they were administrators held a master’s degree and another 11.1% held a doctorate.

Table 4.12. Comparison of Education and Experience of Administrators of Montessori Programs

Education	Panel		Survey	
	Frequency	Percent	Frequency	Percent
Bachelor's Degree	0	0.0	5	27.8
Master's Degree	2	66.7	9	50.0
Doctoral Degree	1	33.3	2	11.1
Other	0	0.0	2	11.1
Total	3	100.0	18	100.0
Education Affiliation				
AMS	1	33.3	8	44.4
AMI	1	33.3	3	16.7
MEPI	1	33.3	1	5.6
None	0	0.0	6	33.3
Total	3	100.0	18	100.0
Years of Experience				
0-3 years	2	66.7	10	55.6
4-7 years	0	0.0	2	11.1
8-11 years	0	0.0	3	16.7
12-15 years	0	0.0	1	5.6
More than 15 years	1	33.3	2	11.1
Total	3	100.0	18	100.0

All administrators on the Panel were Montessori trained as a teacher or administrator, one from an AMS program, one AMI, and one MEPI. Thirty-three percent of Survey administrators had no formal Montessori training. The remainder were trained primarily in AMS programs ($n = 8, 44.4\%$). As indicated in Table 4.11 all three Panel administrators held a Montessori credential, but only one held a Montessori administrator credential.

Two (11.1%) of the Survey administrators indicated a Montessori administrator's credential. This is to be expected when reviewing the answer to the next question, the summary in the second half of Table 4.13 that indicates 78.8% of administrators in the survey were from public Montessori programs and nearly half of those administrators were from school within school or magnet programs, where Montessori is only one part of the school education program.

In contrast, two of the administrators from the Panel group were from a public, non-charter Montessori schools and the third from a private school. In both cases, administrators would be overseeing only Montessori education as opposed to a mix of Montessori and traditional education.

Table 4.13. Comparison of Montessori Certification and Practice Setting in Which Administrators Work

Montessori Certification	Panel		Survey	
	Frequency	Percent	Frequency	Percent
Administration	1	33.3	2	11.1
Early Childhood	2	66.7	3	16.7
Elementary 1	0	0.0	1	5.6
Elementary 1 and 2	0	0.0	1	5.6
Early Childhood and Administration	0	0.0	1	5.6
Early Childhood and Elementary 1	0	0.0	1	5.6
Early childhood, Elementary 1 and 2, and Administration	0	0.0	1	5.6
Not Certified	0	0.0	8	44.4
Total	3	100.0	18	100.0
Practice Setting				
Private Montessori school	1	33.3	4	22.2
Public Montessori classroom within a traditional school	0	0.0	5	27.8
Public Montessori charter school	0	0.0	2	11.1
Public Montessori school (non-charter)	2	66.7	4	22.2
Public Montessori magnet program	0	0.0	2	11.1
Other	0	0.0	1	5.6
Total	3	100.0	18	100.0

Overall, the Panel was a closer representation of public and private practice settings and affiliations in South Carolina but was biased toward rural and suburban schools and the midland

region of the state. The Survey group more closely aligned with the geographic regions of the state but was heavily biased toward public Montessori classrooms with the majority of ECMTs holding a state license in addition to a Montessori credential. Both groups were primarily white females, but without published data on the demographics of the profession it can only be speculated as to whether that is an accurate representation. There were a small number of minority participants represented on the Panel and in the larger Survey group.

The next step in the analysis was to analyze the ratings, combine the two scales (Criticality and Frequency) to create a test blueprint, and then compare the end results from the two groups. The following section will present the results from these analyses.

Analyzing individual tasks. Median, minimum and maximum ratings for Criticality and Frequency ratings are reported for the two groups in Tables 4.14 and 4.15 respectively. The task id (i.e., D1T1) aligns with the content outline presented in Table 4.1. Individual frequency tables reporting Panel and Survey group responses on each scale (Criticality and Frequency) for each task are included in Appendix Q. The b parameters with their standard errors generated from the RRSN using Winsteps software (Linacre, 2016a) are also reported in Tables 4.14 and 4.15.

Recall that the b parameter is the item measure that indicates the difficulty of an item, in this case, how difficult a task is to endorse. The Rasch scale is somewhat arbitrary and any number of units, or logits (i.e., -3 to +3), means and standard deviations may be selected. The default is a mean of 0 and standard deviation of 1, which was used for this study. For example, $b = 0$ would typically be an item of medium difficulty with higher measures being harder and negative measures being easier. The literal meaning of $b = 0$ is that a person who's logit score is at 0 would have a 50% probability of selecting a category whose threshold is located at that point on the scale. In order to avoid confusion in reviewing these parameters, all b parameters have

been multiplied by -1 so that the negative values correspond approximately with lower response options (i.e., 0, 1), and higher values correspond approximately with higher response options (i.e., 4, 5).

Table 4.14. Analysis of Criticality Responses

	Panel (<i>n</i> = 16)					Survey (<i>n</i> = 79)				
	Median	Min	Max	<i>b</i> *	<i>SE</i>	Median	Min	Max	<i>b</i> *	<i>SE</i>
D1T1	5.0	4.0	5.0	1.12	0.52	5.0	0	5	0.36	0.19
D1T2	5.0	3.0	5.0	0.87	0.48	5.0	0	5	0.25	0.19
D2T1	5.0	4.0	5.0	1.77	0.63	5.0	0	5	0.43	0.19
D2T2	5.0	3.0	5.0	0.28	0.42	4.0	0	5	-0.21	0.17
D2T3	4.0	1.0	5.0	-1.51	0.33	4.0	0	5	-1.13	0.16
D2T4	4.0	3.0	5.0	-0.33	0.37	4.0	0	5	-0.48	0.17
D2T5	4.0	3.0	5.0	-0.19	0.38	4.0	0	5	-0.02	0.18
D2T6	4.0	1.0	5.0	-0.59	0.36	4.0	0	5	-0.12	0.18
D3T1	4.5	3.0	5.0	0.11	0.4	4.0	1	5	-0.21	0.17
D3T2	5.0	4.0	5.0	1.41	0.56	5.0	1	5	0.7	0.20
D3T3	5.0	3.0	5.0	1.77	0.63	5.0	1	5	0.39	0.19
D3T4	4.0	2.0	5.0	-1.07	0.34	4.0	1	5	-0.27	0.17
D3T5	4.0	2.0	5.0	-1.07	0.34	4.0	1	5	-0.33	0.17
D3T6	5.0	2.0	5.0	-0.46	0.36	4.0	1	5	0.05	0.18
D3T7	4.0	2.0	5.0	-0.59	0.36	4.0	1	5	-0.3	0.17
D4T1	5.0	2.0	5.0	-0.33	0.37	4.0	1	5	0.18	0.18
D4T2	5.0	2.0	5.0	-0.19	0.38	4.0	1	5	0.28	0.19
D5T1	4.0	2.0	5.0	-1.29	0.33	4.0	1	5	-0.33	0.17
D5T2	5.0	2.0	5.0	0.28	0.42	5.0	1	0	0.78	0.20

*b**(-1) applied to align with order of responses

Table 4.15. Analysis of Frequency Responses

	Panel					Survey				
	Median	Min	Max	<i>b</i> *	<i>SE</i>	Median	Min	Max	<i>b</i> *	<i>SE</i>
D1T1	5.00	4	5	2.27	0.99	5.00	1	5	0.52	0.24
D1T2	4.00	3	5	-0.33	0.29	5.00	1	5	-0.73	0.19
D2T1	4.00	1	5	-0.84	0.26	5.00	1	5	0.06	0.22
D2T2	3.50	1	5	-0.98	0.26	5.00	1	5	0.26	0.22
D2T3	2.50	1	5	-1.8	0.27	4.00	1	5	-1.86	0.18
D2T4	5.00	3	5	0.94	0.49	5.00	1	5	-0.08	0.21
D2T5	3.50	1	5	-1.25	0.26	5.00	1	5	0.16	0.22
D2T6	4.00	2	5	-0.56	0.27	5.00	1	5	-0.21	0.21
D3T1	5.00	4	5	0.94	0.49	5.00	1	5	0.63	0.24
D3T2	5.00	2	5	0.04	0.32	5.00	2	5	0.63	0.24
D3T3	4.00	2	5	-0.91	0.26	5.00	1	5	0.21	0.22
D3T4	5.00	4	5	0.55	0.40	5.00	1	5	0.36	0.23
D3T5	5.00	3	5	0.15	0.34	5.00	1	5	0.11	0.22
D3T6	3.50	2	5	-0.91	0.26	5.00	2	5	0.06	0.22
D3T7	4.50	2	5	-0.15	0.30	5.00	1	5	0.21	0.22
D4T1	5.00	4	5	1.21	0.56	5.00	0	5	-0.16	0.21
D4T2	5.00	2	5	0.27	0.35	5.00	0	5	0.06	0.22
D5T1	3.00	3	5	-0.91	0.26	4.00	0	5	-0.99	0.19
D5T2	5.00	4	5	2.27	0.99	5.00	1	5	0.75	0.25

b*(-1) applied to correct order

Notice that the Rasch standard errors from the Panel responses are more than twice those of the Survey responses. This is largely due to the small sample. As with any measurement model or statistical test, minimum sample size may vary depending on the data. Modern Rasch measurement was founded on research conducted by Wright using 35 students and 18 test items (Linacre, 2016b; Wright & Panchapakesan, 1969), thus designed with small samples in mind. Linacre (2016b) provided guidelines for minimum sample sizes in a Rasch analysis. Sixteen is the absolute minimum mentioned with a 95% confidence interval that is plus or minus 1 logit. The standard errors for both scales are all within this range.

Analyzing the measurement scales using RRSM. A benefit of using the RRSM is the scale structural information that is provided by Winsteps (Linacre, 2016a) in support of evaluating quality. In this analysis, first the individual scales were evaluated using item response graphs followed by review of reliability and fit indexes. Next, responses were combined using formulas presented in Chapter 3 and analyzed using the same graphs and indexes as the individual scales.

Figure 4.5 displays the category probability curves from Winsteps (Linacre, 2016a). As the legend at the bottom of the graph denotes, each curve aligns with a point on the original rating scale. A properly performing response scale should have distinct peaks, with the number of peaks depending on the number of scale points, and one peak being higher than the others indicating a higher response probability (Bond & Fox, 2007). Such a distribution provides clear thresholds, or locations where two curves intersect, indicating the point on the logit scale where the probability of selecting either category is at 0.5. For example, in the first graph in Figure 4.5, the point where category 1 and 2 intersect is at approximately -2.5, indicating that a person with a fairly weak opinion of a task (-2.5 on the scale) would have a 50% probability of selecting one of the two categories. Criticality is performing fairly well for both groups; however the lowest category is missing for the panel group due to the lack of a zero response in the dataset. All of the thresholds are a little close together around the middle of the scale, but is sufficient for the purposes of this study.

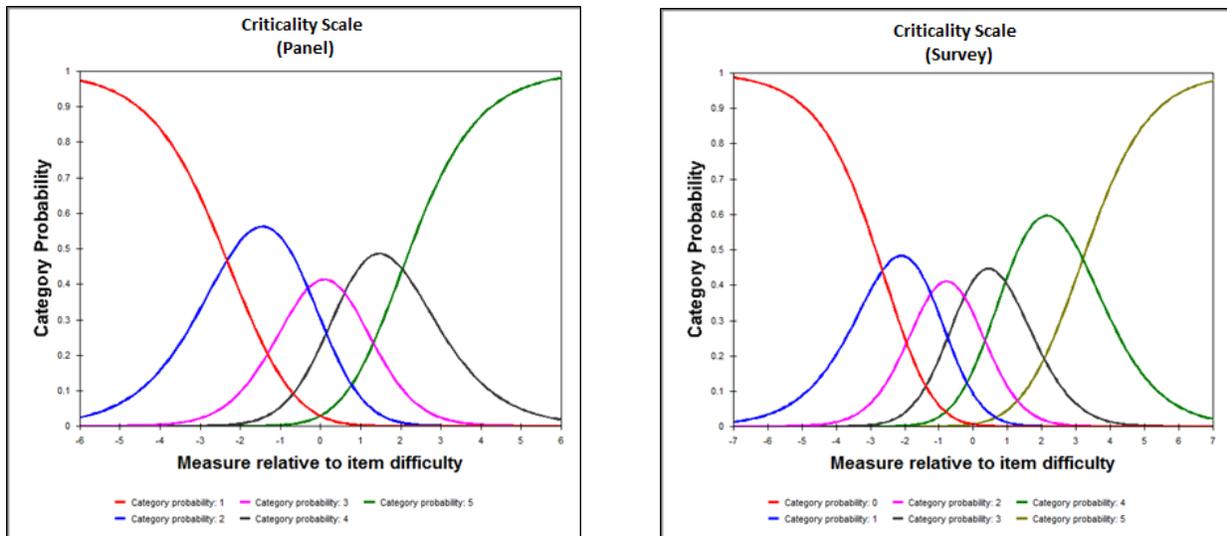


Figure 4.5. Comparison of criticality rating scales.

The frequency scale is performing similarly for the survey group (Figure 4.6), however Category 4 is not as clearly delineated for the panel group as is ideal. It indicates that the frequency of selecting category 4 in the panel group was fairly low. Important to note is the presence of ordered categories that is expected with Likert-type rating scales. Disordered category probabilities would have been an indication of an issue with the response format.

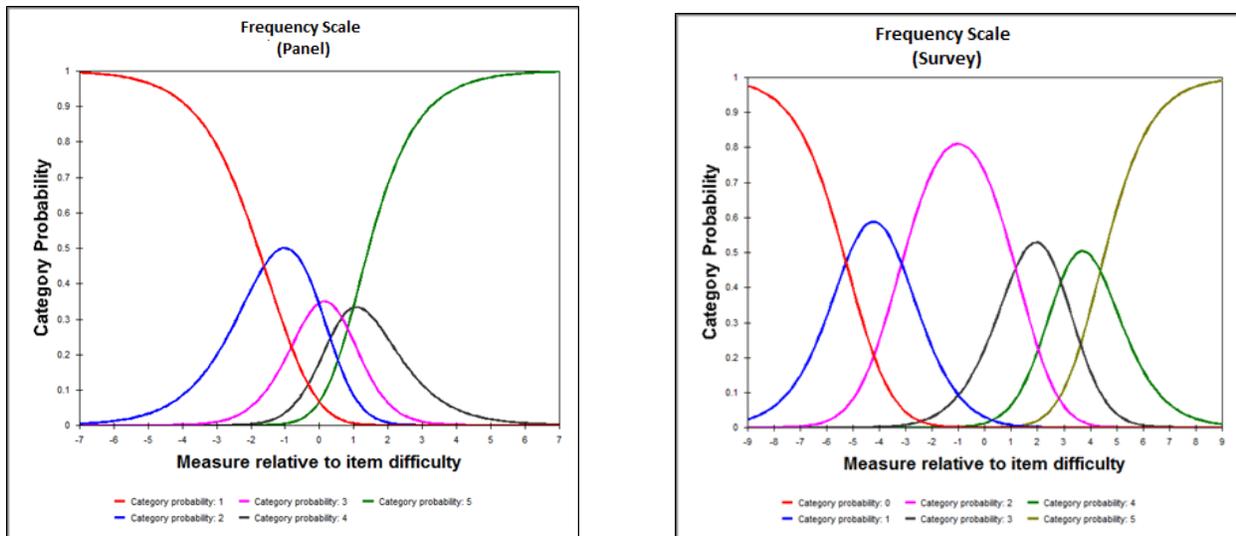


Figure 4.6. Comparison of frequency rating scales.

The next step of the analysis was to review Rasch fit statistics and reliability, presented in Table 4.16. Cronbach's alpha was 0.97 for Criticality and Frequency with the Survey group and lower for the Panel group (0.92 and 0.76), particularly on the Frequency scale. Small sample likely impacted the Frequency scale reliability.

Item reliability and separation are estimates of scale quality related to the RRSM analysis. Item separation is an indicator of performance of the response structure. Separation should be small (< 3) (Linacre, 2016b). High separation and low item reliability would be indicative that the sample size was too small to adequately estimate the order of responses. Measures between 2 and 3 are not informative if the goal is to construct a measurement scale across items (2016b), but these values are sufficient for the purposes of this study.

Table 4.16. Reliability of Criticality and Frequency Scales

	Criticality		Frequency	
	Panel (<i>n</i> = 16)	Survey (<i>n</i> = 79)	Panel (<i>n</i> = 16)	Survey (<i>n</i> = 79)
Alpha	0.92	0.97	0.76	0.97
Item Reliability	0.78	0.81	0.82	0.86
Separation	1.88	2.08	2.11	2.47
Model <i>SE</i>	0.47	0.18	0.4	0.22
Fit at mean	0.98/0.99	1.01/.96	0.99/1.00	0.99/0.97
Fit at max	2.01/2.16	1.6/1.58	1.63/2.15	1.69/1.78
Fit at min	0.45/0.47	0.64/0.56	0.39/0.39	0.58/0.46

Item reliability indicates the consistency with which items are located on the logit scale and is influenced by sample size and range of difficulty of the items (*b* parameters). These values are lower for the panel group on Criticality due to small sample and limited range of responses. Similar to the review of *b* parameters, the Rasch standard errors are approximately twice that of the survey group, but are within one half logit of the measure, thus within acceptable parameters.

Item fit statistics, presented in the last three rows of Table 16, are indicative of how well the data fit the RRSM. Indexes are expected to be within the range 0.5 to 1.5 logits (Linacre, 2016b). Two indexes are reported, the first being Infit Mean Square and the second Outfit Mean Square. Infit measures how items are performing for the targeted theta (logit) value. Outfit indicates unpredictable responses or outliers where the variability in scores is considered noise. The fit of tasks to the model at the mean *b* value are all indicating good fit (close to 1). At the extremes of the scale indexes fall out of range. For minimum *b* parameters, the indexes are only slightly out of range. The most concerning are the Panel outfit statistics at the maximum values of *b*, which are likely due to the very small number of responses at the low end of the scale.

Analyzing combined scales. A formulation for combining scores and creating blueprint weights was described in Chapter 3. Table 4.17 provides the medians, minimum and maximum

responses, reversed b parameters, and percent weights for each task. Individual frequency tables for the combined results are presented in Appendix R. The percent weights reported in Table 4.17 indicate the proportion of a test that would cover the knowledge and skills related to that specific task. The lowest rated (D2T3) and highest rated (D5T2) tasks for both groups are the same, although the Panel group rated a second task as highest (D1T1). Similar to previous analyses, standard errors are larger for the panel group.

Table 4.17. Analysis of Combined Responses

	Panel						Survey					
	Med	Min	Max	<i>b*</i>	<i>SE</i>	PWT	Med	Min	Max	<i>b*</i>	<i>SE</i>	PWT
D1T1	6.00	5.48	6.00	1.73	0.50	8.14	5.48	2.45	6.00	0.46	0.21	6.11
D1T2	5.48	4.47	6.00	0.38	0.40	5.90	5.48	2.45	6.00	-0.13	0.20	5.02
D2T1	5.24	3.46	6.00	0.08	0.39	5.40	5.48	2.45	6.00	0.42	0.21	6.04
D2T2	4.95	3.46	6.00	-0.62	0.37	4.23	5.48	2.45	6.00	-0.25	0.20	4.80
D2T3	3.87	2.45	6.00	-2.16	0.36	1.67	4.47	2.45	6.00	-1.85	0.18	1.85
D2T4	5.48	4.47	6.00	0.71	0.41	6.44	5.48	2.45	6.00	-0.44	0.20	4.45
D2T5	4.47	3.16	6.00	-1.40	0.36	2.93	5.48	2.45	6.00	-0.17	0.20	4.95
D2T6	5.24	3.16	6.00	-0.76	0.36	4.00	5.48	2.45	6.00	-0.25	0.20	4.80
D3T1	5.48	4.90	6.00	0.89	0.43	6.74	5.48	2.45	6.00	0.00	0.20	5.27
D3T2	5.74	3.87	6.00	0.89	0.43	6.74	5.48	3.00	6.00	1.02	0.22	7.15
D3T3	5.48	3.87	6.00	-0.21	0.38	4.91	5.48	2.45	6.00	0.33	0.21	5.87
D3T4	5.24	3.87	6.00	-0.07	0.38	5.15	5.48	2.45	6.00	0.00	0.20	5.27
D3T5	5.24	3.46	6.00	-0.07	0.38	5.15	5.48	2.00	6.00	-0.04	0.20	5.19
D3T6	5.19	3.00	6.00	-0.76	0.36	4.00	5.48	2.45	6.00	0.16	0.21	5.56
D3T7	5.48	3.46	6.00	-0.21	0.38	4.91	5.48	2.45	6.00	-0.17	0.20	4.95
D4T1	6.00	4.24	6.00	0.71	0.41	6.44	5.48	2.24	6.00	0.16	0.21	5.56
D4T2	5.74	3.87	6.00	0.55	0.41	6.18	5.48	2.24	6.00	0.29	0.21	5.80
D5T1	4.47	3.46	6.00	-1.40	0.36	2.93	5.00	2.24	6.00	-0.67	0.19	4.03
D5T2	6.00	3.87	6.00	1.73	0.50	8.14	5.48	3.00	6.00	1.11	0.22	7.32

b(-1)* applied to correct order

One of the arguments in favor of the multiplicative model selected was that it retained the ordinal structure of the scales, as demonstrated in the two graphs of Figure 7. As with the

previous review of category structure for the individual scales, these images indicate that the combined scales maintain the ordinal structure of the data. Recall in the formula used to combine the two scales, one was added to Criticality in order to prevent zero weights for any task. The combined scale is, thus, one to six rather than the original 0-5. As there were no combined responses below 2, category one does not appear in the analysis.

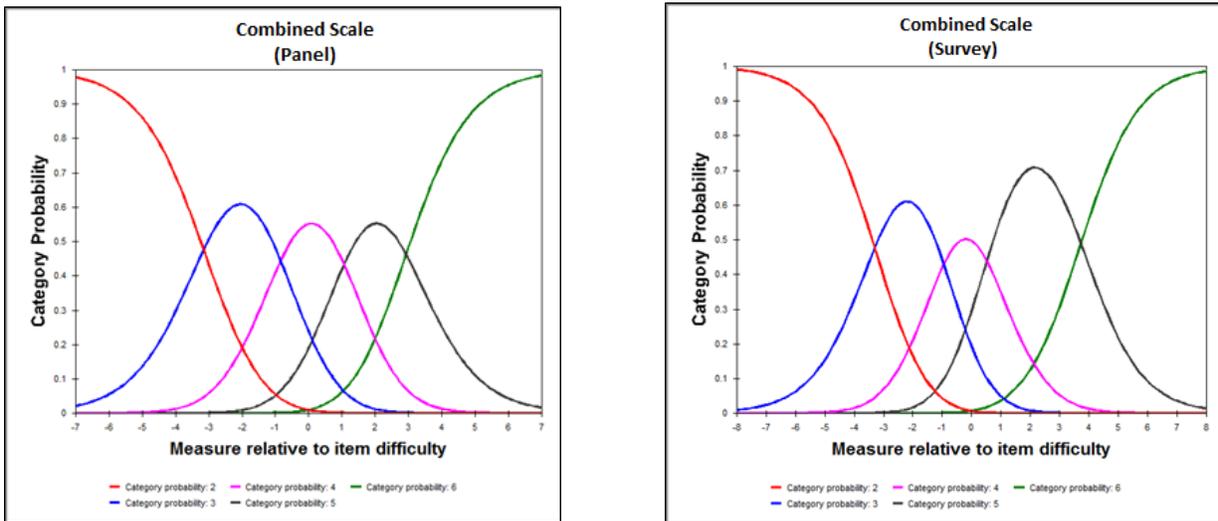


Figure 4.7. Comparison of combined rating scales.

Reliability and fit statistics are presented in Table 4.18. Values are mostly within acceptable ranges. Although the fit statistics at the minimum categories are below 0.5, they are not concerning. These indicate an overfit, or response patterns that are overly predictable such as when the same one or two response categories are used for a task (Linacre, 2016b). It does not negatively impact the analysis for purpose of this study. The fit at the maximum values are only slightly out of range according to Linacre's (2016b) guidelines.

Table 4.18. Reliability Comparison

	Panel	Survey
<i>KR20</i>	0.9	0.96
Item Reliability	0.82	0.88
Separation	2.15	2.72
Model <i>SE</i>	0.40	0.20
Fit at mean	0.98/0.98	1.0/0.97
Fit at max	1.60/1.46	1.60/1.68
Fit at Min	0.25/0.26	0.63/0.60

In addition to the fit statistics and reliability indexes, Q-Q plots were reviewed to evaluate normality for the purpose of using correlation to compare groups (Harwell & Gatti, 2001). Plots for the two samples are compared in Figure 4.8. Panel data follows the fit line fairly well, although the few number of data points may make diversions from the line difficult to detect. The Survey data appear to be forming a slight s curve, particularly at the low end where there was less data. Clear interpretation of data normality is difficult with these small sample sizes.

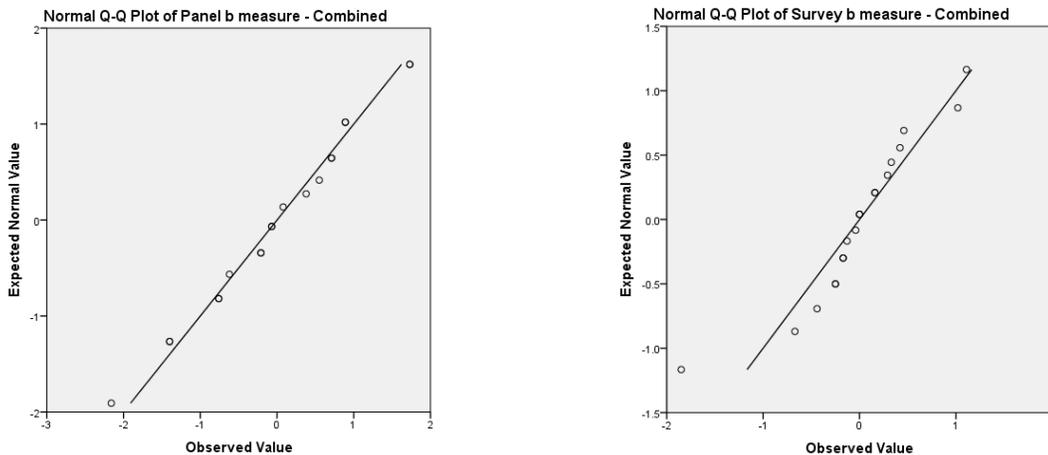


Figure 4.8. Q-Q Plots using panel and survey *b* parameters to test for normality.

Results reported in this section have been an evaluation of the two rating scales used on the questionnaire, Criticality and Frequency, and the combined scale. Analyses indicate that scales met standards for reliability and category structure based on IRT analysis. Although the multiplicative model has potential limitations for combining scales, the model selected provided an acceptable option to support the use of the RRSM that assumes data is at a minimum ordinal. Finally, reliability and fit indexes supported the use of the model and provided some evidence that the data meet assumptions of normality, although small samples preclude a definitive answer. The next section will describe the statistical comparisons of median ratings between Panel and Survey groups using Mann-Whitney U test of rank order, correlation of b parameters, and comparison of proportions of test between groups.

Group comparisons. Methods for comparisons were determined in part by previous research. Maurer and Tross (2000) and Tannenbaum and Wesley (1993) both compared panel ratings with the larger survey population ratings using a Pearson Product Moment Correlation, Intraclass correlation, and a dichotomous index that required establishing a cut point on the combined rating scale. As mentioned early, the two parametric tests used assume the data is normally distributed. This makes inferences drawn from these analyses suspect when using data from rating scales that produce ordinal data and is problematic with very small sample sizes, particularly an issue with the Panel data, both in previous and current studies.

Although the data in the current study were transformed using RRSM and there is some evidence that the data are not normally distributed and have interval properties, only the correlation analysis will be replicated in this study in addition to non-parametric analyses used to support results. First, ratings between the two groups were compared using a non-parametric Mann-Whitney U (MWU) test of rank order, which is appropriate for use with ordinal data.

Second, correlations were estimated using the Rasch calibrated b measures rather than mean ratings that were used in the two referenced studies. The dichotomous index used in those studies was not applied here as the median ratings were all fairly high and would not be informative for this study.

Results of Mann-Whitney U. MWU estimates and compares a mean rank for each group to test the null hypothesis that two groups originate from the same population (Field, Miles, & Field, 2012; Howell, 2002). This can also be viewed as testing the null hypothesis that the ranks in one group are not “systematically higher or lower than the other” (Corder & Foreman, 2014, p. 72). It is the non-parametric equivalent to comparing two groups using a t-test, thus, allowing an analysis using the original combined scores from the survey. The evidence provided by this test aid in determining whether the panel is representative of the population and where in the content domain there are significant differences in opinion between the two groups.

Data met the following assumptions (Lund & Lund, 2013): data was at least interval level, independence of observations, and groups (Panel and Survey) were independent. In addition, when there are ties, meaning that multiple scores are the same, a different algebraic formula must be used to adjust for these (Corder & Foreman, 2014). SPSS automatically detects these ties and uses the correct formula for estimating the test statistic.

MWU also assumes normal distributions, particularly that distributions being compared have similar kurtosis, which can be analyzed using population pyramids generated by SPSS (Figure 4.9 and Appendix S). Distributions were also evaluated using Kolmogorov-Smirnov test, which compared the distribution of each task to an empirically derived normal distribution. Given small samples and restriction of range, particularly in the Panel group, distributions were not normally distributed for many of the items, including visual comparison of shapes across

groups. With non-normal distributions, we can infer whether one group's mean rank is higher or lower than the other and can compare these mean ranks using SPSS, but cannot compare medians (2013). Even with comparing mean ranks, statistical significances should be interpreted with caution.

To test the null hypothesis, each set of responses - one set from the Panel and one set from the Survey- are ordered according to magnitude and ranks assigned. Ranks within each group are then summed to estimate R in the following formula:

$$U = n_1 n_2 + \frac{n_2(n_2 + 1)}{2} - \sum_{i=n_1+1}^{n_2} R_i$$

where n is the sample size of each group and R is the ranks of the sample. A standardized test statistic is calculated and significance determined from the U distribution. A mean rank is also estimated for each group that adjusts for sample size.

This test was conducted using SPSS (IBM Corp., 2013) across all 19 tasks using the combined scale results for the 16 panel and 79 survey respondents ($n = 95$). Table 4.19 provides a summary of the results. Mean ranks can be compared between the panel and survey groups in the first two columns of reported data. The unstandardized and standardized test statistics (U), standard error, and asymptotic significance values are also reported. At a significance level of $p < .05$, two tasks have a statistically significant difference between mean ranks. At a significance level $p < .01$, the null hypothesis is retained for all tasks. As mentioned above, statistical significance provides some insight into where there may be stronger differences, but must be interpreted with caution due to non-normal distributions.

Table 4.19. Summary Mann-Whitney U test of Rank Order Between Independent Groups

Task	<i>n</i>	Panel Mean Rank	Survey Mean Rank	Unstd <i>U</i>	Std <i>U</i>	<i>SE</i>	Sig. (2-sided)
D1T1	95	60.00	45.57	824	2.041	94.053	.041*
D1T2	95	51.34	47.32	685.5	.545	98.081	.585
D2T1	95	43.94	48.82	567.0	-.677	96.050	.499
D2T2	95	41.5	49.32	528	-1.067	97.502	.286
D2T3	95	38.81	50.27	453	-1.796	99.685	.073
D2T4	95	55.97	46.39	759.5	1.299	98.177	.194
D2T5	95	34.12	50.81	410	-2.297	96.648	.022*
D2T6	95	41.62	49.29	530	-1.044	97.689	.296
D3T1	95	55.75	46.43	756.0	1.285	96.507	.199
D3T2	95	50.72	47.45	675.5	.464	93.725	.643
D3T3	95	43.78	48.85	564.5	.698	96.696	.485
D3T4	95	46.16	48.37	602.5	-.300	98.258	.764
D3T5	95	48.53	47.89	640.5	.086	98.364	.931
D3T6	95	38.75	49.87	484	-1.513	97.850	.130
D3T7	95	47.94	48.01	631	-0.010	98.110	.992
D4T1	95	59.06	46.37	761	1.322	97.584	.186
D4T2	95	50.44	47.51	671	.401	97.216	.688
D5T1	95	38.25	49.97	476	-1.575	99.034	.115
D5T2	95	55.88	46.41	758	1.350	93.304	.177

*significant at $p < .05$

In addition to the test statistics and p -values, review of bar graphs provides a visual comparison of the two distributions for each task. Figure 9 is a comparison of response distributions between Survey and Panel groups for the two tasks where the null hypothesis is

rejected at the .05 level (D1T1 and D2T5). For task D1T1, the Panel responses were all high (4 and 5), whereas the Survey responses were more varied. In the second example (D2T3) the distributions were more similar, however, there were fewer high values (4 and 5) selected by Panel participants producing a much lower mean rank.

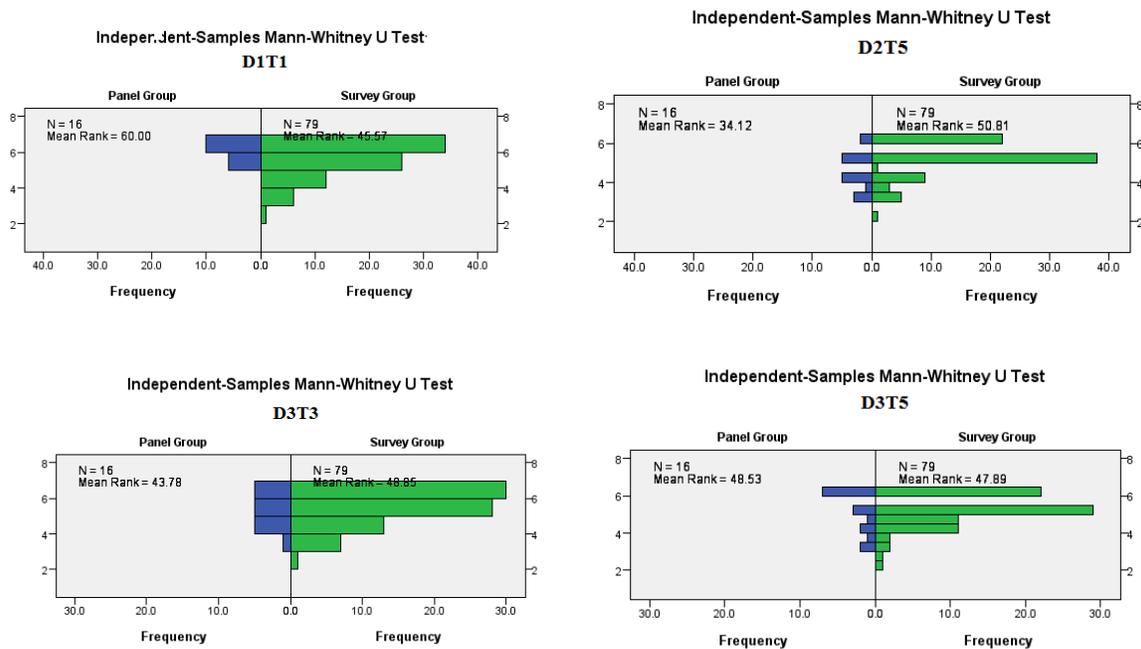


Figure 4.9. Comparison between survey and panel responses across mean scores.

The final two examples in Figure 9 present examples where distributions produced a more similar mean rank. Task D3T5 has a more widely spread distribution similar to D2T5 above, but with a higher concentration of scores at the top of the scale producing a higher mean rank for both groups.

All 19 tasks were visually analyzed (see Appendix S for the remaining 15 graphs). Data were also analyzed using only responses from teachers. The null hypothesis was retained for all 19 tasks under this condition.

Test of Proportions. Table 4.20 contains the final blueprint for a hypothetical ECMT certification test that could be developed and administered in South Carolina. The primary research question for this study essentially asks whether the blueprint developed by a panel of experts would be comparable to the blueprint developed based on the results of a survey of experts from the broader population. The mean ratings that were scaled to create these weights were compared in the MWU analysis, indicating statistical differences ($p < 0.05$) with a two tailed test for only two tasks (D1T2, and D2T5). Imagine two 100 item tests developed, one based on Panel results and the other based on Survey results. Test forms would differ by two items for the two aforementioned tasks. In fact, the two blueprints never vary more than 2 items at the task level. As will be seen in the test of proportions analysis in the next figure, overall differences are not statistically significant.

Table 4.20. Comparison Between Groups of Proportion of Test Content

	Percent of Test (Panel)	Percent of Test (Survey)	Difference on a 100 item test
Domain 1	14.04	11.14	3
Task 1.1	<u>8.14</u>	<u>6.11</u>	<u>2</u>
Task 1.2	5.90	5.02	1
Domain 2	24.66	26.90	2
Task 2.1	5.40	6.04	1
Task 2.2	4.23	4.80	1
Task 2.3	1.67	1.85	0
Task 2.4	6.44	4.45	2
Task 2.5	<u>2.93</u>	<u>4.95</u>	<u>2</u>
Task 2.6	4.00	4.80	1
Domain 3	37.60	39.26	2
Task 3.1	6.74	5.27	1
Task 3.2	6.74	7.15	0
Task 3.3	4.91	5.87	1
Task 3.4	5.15	5.27	0
Task 3.5	5.15	5.19	0
Task 3.6	4.00	5.56	2
Task 3.7	4.91	4.95	0
Domain 4	12.62	11.36	1
Task 4.1	6.44	5.56	1
Task 4.2	6.18	5.80	0
Domain 5	11.07	11.34	0
Task 5.1	2.93	4.03	1
Task 5.2	8.14	7.32	1

A chi square test of association was used to test whether the proportions were from the same population.

$$\sum_{r=1}^R \sum_{c=1}^C \frac{n_c}{100} * \frac{(P_{rc} - 100\hat{\pi}_r)^2}{100\hat{\pi}_r}$$

where rows for this analysis were the task percent weights estimated in the blueprint ($n = 19$), and columns were group values ($n = 2$). P_{rc} was proportion of the test covered by that task (row)

when divided by the total of all proportions, in this case 100 since the weights total 100%. In the final portion of the formula, $\hat{\pi}$ was the expected proportion for each row.

Analysis of the Chi-Square results indicated a non-significant statistic ($\chi^2 = 1.11, p = 1.0, df = 18$). The overall differences between the two groups were not significant. Figure 10 is a graphical display of this analysis. Distributions of proportions of the test are indicated for both groups compared to an average. Evaluation of the 95% confidence interval for each task indicates where there was more variability in task weights between the groups. Similar to the results of the MWU analysis, D1T1 and D2T5 show greater differences. Domain 2 Task 4 also has a large confidence interval. Although not statically significant ($p = 0.194$), the difference between mean rank values in the MWU for that task was nearly 10 points. Similarities are also more apparent in this graphical comparison, particularly where there was almost complete agreement such as D2T3. Both groups agreed this was the least important task in the content domain.

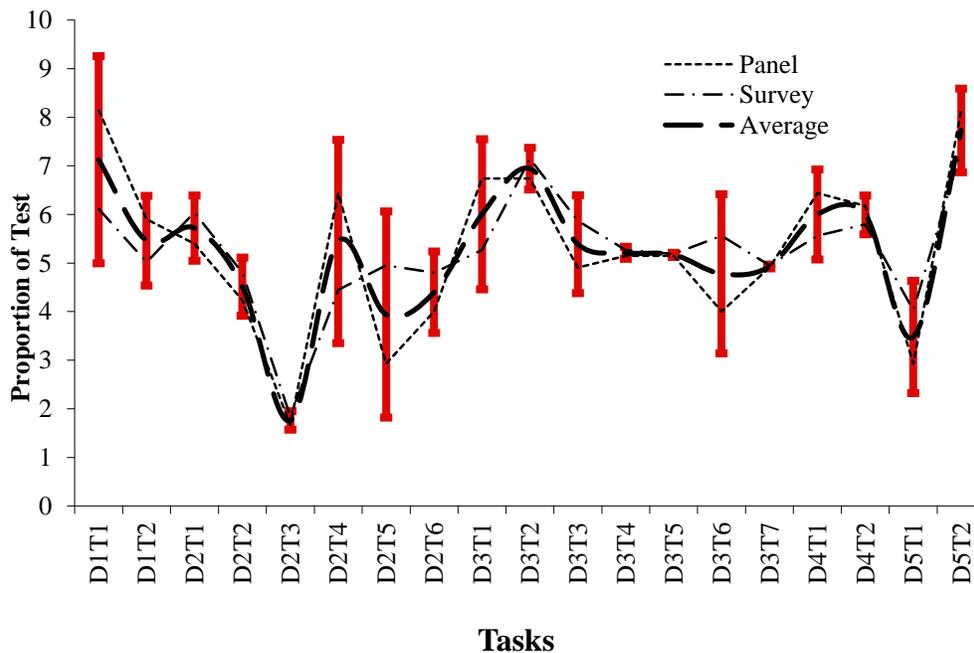
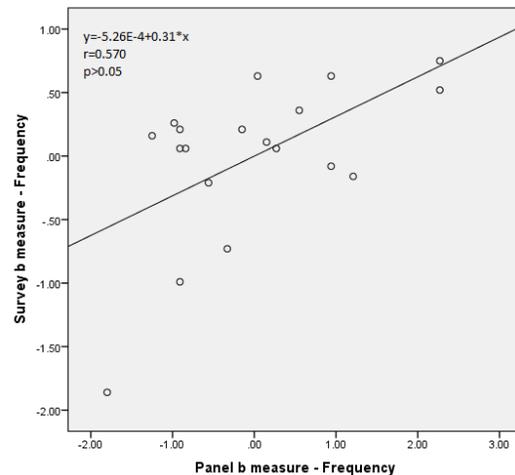
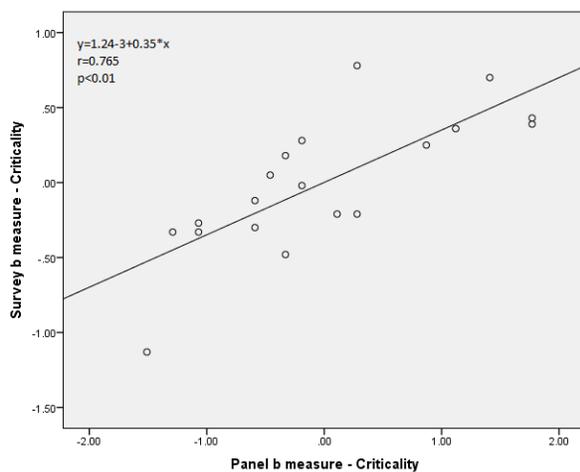


Figure 4.10. Proportion of test comparison between panel and survey responses.

The results of the Mann-Whitney U test provided evidence that the responses to the two groups were from the same population across most of the tasks. The test of proportions also indicated agreement over most of the content domain and a non-significant chi-square statistic that provides further evidence of the association between the Panel and Survey groups. Next, a correlation analysis was conducted to examine the strength of the relationship.

Correlation. Correlation analysis was conducted to compare the strength of the relationship between the two groups and to compare to previous research. Comparison was made using the reversed *b* parameters reported in Table 4.14 in order to meet the assumption that variables are continuous. Although the focus of this analysis is comparison between groups on the Combined scale in order to answer the research questions, correlations between the individual scales is reported as well.

The relationship between group responses was first compared graphically in Figure 11. Responses to the Combined scale have a fairly strong, positive correlation. Criticality alone still suggests a fairly strong, positive correlation. However, correlation on the Frequency scale is lower and does not appear to meet assumptions of linearity and homoscedasticity.



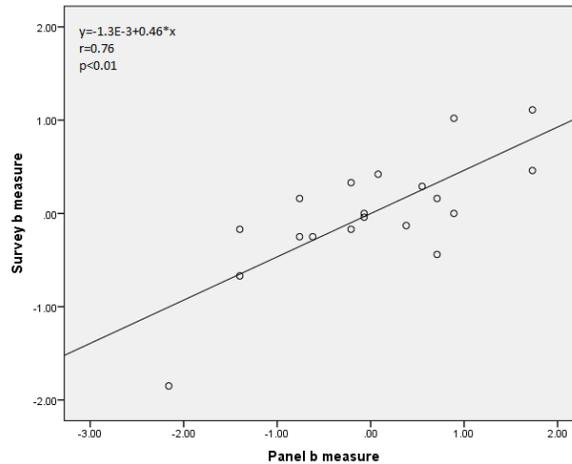


Figure 4.11. Comparison of mean b parameters for combined, criticality and frequency scales.

Given possible violations of assumptions for Pearson's r with the small sample sizes and non-normal distributions, Spearman's correlation coefficient and Kendall's tau were also calculated (Field et al., 2012). Results of all three tests are presented in Table 4.21.

Table 4.21. Summary of Correlation Analysis for All Three Response Scales

Rating Scale	Pearson r	Spearman	Tau
Criticality	0.765**	0.820**	0.653**
Frequency	0.570*	0.425	0.295
Combined	0.760**	0.683**	0.544**

*significant at $p < .05$

**significant at $p < .01$

Pearson's r on the Combined scale was 0.760, significant at the 0.01 level. The correlations reported in the two previous studies were 0.82, 0.86, and 0.84 (Maurer & Tross, 2000; Tannenbaum & Wesley, 1993), but these values are suspect due to possible violations of

assumptions. Spearman's coefficient and Kendall's Tau, while lower, still suggest a moderate correlation on the combined scale.

Pearson's r when comparing the two group's responses to the Criticality scale is nearly the same as the combined scores, but only a moderate correlation between groups on the Frequency scale. Correlations are very small between groups on the Frequency scale when applying Spearman or Kendall's Tau. Reflect back to the median frequency ratings reported in Table 4.15. There was much more variability in ratings within the panel group than the survey group.

Additional tasks from Survey participants. In addition to rating tasks in each domain, Survey participants were asked to list any additional essential tasks that pertained to that domain that were not included in the list. The goal of this in a typical practice analysis survey is to identify any possible gaps in the blueprint. However, it is difficult for those responding to a survey to provide informative additions given that they have not been exposed to training that the panel of experts has. Thus, most of the comments are not actually tasks but general comments or elements of what a teacher does that are encompassed in the existing tasks or would be included in the KSAs.

Seventeen comments were made from seven of the 79 survey participants. A complete list of comments included in Appendix T, and summarized with researcher comments in Appendix U. Examples of some of the general comments about quality teaching are "being consistent is very important," and "there should be a wonderful communication relationship between the student and the teacher." Other comments were related to existing tasks, such as one regarding classroom/behavior management that was included in the first domain before the participant would have seen the tasks related to classroom management in a later domain.

Review conducted by the researcher did not reveal any actual tasks that were not incorporated in the existing list, but in the event that a testing program was to be developed based on this study, the list would be submitted to a SME or SME group for additional review.

Summary

Review of demographics provided evidence that, while somewhat biased toward certain geographic characteristics, the Panel represented affiliations and practice settings of the wider SC Montessori teacher profession. However, the Survey group was heavily biased toward public Montessori programs. Triangulation of results from the three statistical analyses (MWU, chi-square test of association between proportions, correlation) indicated similarities between Panel and Survey group results across most of the blueprint. There may be some areas, particularly two tasks (D1T1 and D2T5), where the blueprint would vary significantly between the two groups.

The following conclusions will discuss these results within the validity framework presented in the Introduction, considering the answer to the final research question regarding impact on validity if only the Panel results were used to establish the test blueprint. Limitations of the study will also be addressed as well as future directions for this research.

CHAPTER 5. DISCUSSION

A three-phase practice analysis study, typical of studies found in related literature over the past 10 years, was conducted to address the question of whether absence of a survey executed in the last phase of the study would negatively impact the strength of a validity argument for a credentialing test. The first phase of the study focused on development of core information and materials by a small committee of subject matter experts, followed by a one day in-person meeting with 13 early childhood Montessori teachers (ECMTs), teacher trainers, and administrators. In this meeting, experts were asked to identify the essential tasks of an entry-level ECMT. Using these tasks as the basis for a questionnaire, a survey was then administered to a broader sample of experts that asked participants to rate each task based on the frequency with which an ECMT performs each task and the risk of negative consequences if an ECMT cannot competently perform these tasks. Results of the first two phases were then reported in Chapter 4 and survey data from Phase III used to develop two versions of a test blueprint, based on each of the Panel and Survey groups. The following discussion focuses on examining the results in relation to the research questions.

Evidence Based on Representation

The overarching research question specifies the condition of a carefully selected panel of subject matter experts and a sample of subject matter experts from the broader target population. Evidence presented, beginning in the Methods description in Chapter 3, included the process used to carefully select subject matter experts (SMEs) for participation on the Advisory Group (AG) and Panel. The specified plan (see Table 3.2) included teachers, teacher trainers, and

administrators from both private and public schools across the Montessori affiliates who had training programs accredited by Montessori Accreditation Council for Teacher Education MACTE in South Carolina (SC). In addition, demographics of SC were considered, although consideration was secondary to practice setting (private/public) and affiliation of the training program when selecting participants given the potential for greater variation in the professional role among these factors. Particularly related to affiliation of training program, there is a general perception that differences in Montessori teaching exist.

The combined Panel/AG group adequately represented most of these facets. Panel participants represented teachers trained in American Montessori Society (AMS), Association Montessori International (AMI), Montessori Education Programs International (MEPI), and Independent (IND) affiliated training programs. Private and public schools were evenly represented, with six public and six private school teachers currently working in classrooms. Although there were more participants (11 of 16) from the Midland region of SC (Columbia and surrounding areas), likely due to the location of the in-person meeting, there was representation from across the state. Urban regions of the state were underrepresented (2 of 12 teachers); however, the panel did include one ECMT from Richland One, even though the district was not officially part of the study.

The survey group sampled was less balanced with regards to practice setting and affiliation, but with more balanced participation across the state. Public teachers and administrators had much stronger representation, with 83.1% of responding ECMTs teaching in public Montessori programs. In addition, the survey sample was more biased toward AMS trained teachers than the Panel group. The panel distribution of affiliations was closer to the distribution of MACTE accredited training programs.

Both groups reflected a primarily white, female population ranging in age from 26 to over 55. Teachers were well educated, with nearly all participants in both groups holding at least a bachelor's degree and many holding higher-level degrees. Survey participants indicated being relatively inexperienced with more than half of participants indicating they had 0-3 or 4-7 years of experience. Teachers in the Panel group were more evenly distributed between newer teachers and those with more experience. Without industry-level data regarding the demographics (i.e., race, gender) of the profession, it is impossible to assess how well either sample represented the overall demographics of the profession.

Distinctions between Panel and Survey groups demonstrated the difference between a carefully selected group and a survey sample, where there is little control over who responds; thus, results of the survey may not adequately represent the profession, particularly when population and/or sample size is small. Studies reviewed in the literature tended to invite large numbers of professionals without acknowledging in the study any possible failings of this large sample method. Few studies reported stratified samples that might address representation more adequately. Future practice analysis studies may benefit from looking more closely at sampling methods that improve sampling across demographics and/or practice settings.

Results of the current study provide some evidence that a carefully selected panel could provide a closer representation of the profession, and a resulting blueprint based on panel ratings alone provide adequate support for validity evidence. Further supporting evidence can be inferred from analysis of rating responses.

Question 1: Final ratings

Table 4.20 in the results provided final blueprint weights reported as a percentage of total test content for each Task and Domain. These weights were calculated by first applying a

multiplicative model to combine raw task ratings from the two rating scales and then separately scaling the combined ratings from 16 panel and 79 survey participants using RRSM. The highest weighted content Domain for both groups was Classroom Management, with 37.6%/39.26 % (panel/survey weightings), and the lowest weighted Domain was Communication and Interaction at 11.07%/11.34%. The next step was to compare task ratings from the two groups in order to answer Question 2.

Question 2: Group comparisons

Comparisons of Panel and Survey group ratings were intended to extend the research conducted by Maurer and Tross (2000) and Tannenbaum and Wesley (1993) to a different set of conditions, expecting to add to the evidence in support of using Panel data alone to create a supportable test blueprint. Similar to previous studies, analysis produced moderate to strong relationships between ratings from Panel and Survey groups. Based on similarities across the three studies, it may be inferred that when the purpose of the survey is to verify that relevant perspectives are represented and the content domain adequately represents practice settings, the panel results may stand on their own as the foundation of a test of competency and as the primary source of validity evidence for a credentialing test. This has certain caveats that will be address in the discussion of validation. First, a comparison of the results was conducted in order to assess the strength of the relationship for this study.

Comparisons between groups included a non-parametric (MWU) test, a test of proportions using a chi-square test of association, and correlation analysis. Non-parametric analysis with the MWU indicated non-significant differences between ratings provided by the two groups across 17 of the 19 tasks with the caveat that the significance test should be interpreted with caution. Significant differences in ratings between Panel and Survey participants were indicated for Task 1 related to presenting lessons in the Instruction Domain (D1T1). The

Panel rated this task higher than the survey group. Opinions related to preparing the environment using Montessori guidelines (D2T5) also differed, with the Survey group viewing this task as significantly more important than the Panel group. Important to note in this comparison is that the relationships are based on the combination of the frequency and criticality responses to create a single indicator of importance. There was stronger disagreement when asked about frequency of performance versus criticality of the task as indicated by both non-parametric analyses and lower correlations. Coupled with the bias in the Survey sample toward public ECMTs and administrators, these differences may reflect differences in the frequency with which public school teachers perform these activities in comparison with private school teachers. Sample sizes were too small to make a meaningful statistical comparison between the two groups with public and private teachers isolated. This may be a topic for future research.

Following the MWU analysis, a test of proportions compared blueprint percent weights between the two groups. Similar to MWU results, D1T1 and D2T5 represented the largest difference in opinion between the two groups. Analysis also demonstrated the points of strongest agreement between groups. D2T3, creating original lessons, was the least important task, and D5T2, establishing partnerships with classroom and school staff, was considered the most important.

Finally, estimated correlation indexes using Rasch parameters for all three response formats (Criticality, Frequency, and Combined) were statistically significant when comparing the two groups using Pearson's r and on the Criticality and Combined responses using Spearman's coefficient (Table 4.21). Kendall's Tau was also estimated due to possible violations of assumptions for r . Correlations were still statistically significant for Criticality and Combined, but on the Frequency scales was very low ($r = 0.29$). The ultimate goal of the comparison was to

examine the similarities between the two blueprints, thus, the significant correlations on the combined scales are indicative of a strong relationship between opinions in the two groups.

Overall, results indicated strong similarities. In addition to the percent weights, Table 4.20 (last column) presented the difference in number of test items when creating a hypothetical test of 100 closed-response (i.e., multiple-choice) items. In this scenario, a test constructed strictly according to this plan would never vary by more than two test questions under each task and only four of the 19 tasks varied by that much. Six tasks would not contain any variation and nine would only vary by one test item. In all cases, task weights for both groups were within a 95% confidence interval. The next question, then, is whether these results are sufficient as the primary source of validity evidence for a credentialing test.

Question 3: Sufficient evidence for validity

First, consider the question of how these results provide content-related evidence for the validity argument in support of test score interpretations that would be generated from a future certification test developed with both the Panel and Survey results combined. Evidence begins with a description and rationale for the chosen methodology. Methods for this study were justified due to the primacy of this method for conducting a practice analysis study as described in the literature review. The method meets requirements set forth in the Standards (AERA, APA, & NCME, 2014) where a test population was identified, test purposes and intended uses were established, qualifications and training of SMEs was documented, and the content domain was clearly defined and evaluated using representative experts. Similarly, method, qualifications and representation of SMEs, and documentation appear to meet the demands set forth in accreditation standards (Institute for Credentialing Excellence, 2010; International Standard Organization Committee on Conformity Assessment, 2012).

Another way to examine the evidence is within the framework of an argument, beginning with the interpretive argument as described earlier (Kane, 2006; M. T. Kane, 2013). The central interpretive argument (IA) for this testing program states: A passing score indicates that an EMCT has demonstrated mastery of the knowledge and skills necessary to competently teach in an early childhood Montessori classroom. Recall that the second type of argument Kane (2013; 2006) described was the validity argument, the purpose of which is to evaluate the IA or “the plausibility of the claims based on the scores” (p.1). The claim of the stated IA specifies that a passing score represents mastery of knowledge and skills. From a content perspective, the focus of this study, this assumes the test content is based on the knowledge and skills required for competent practice of an ECMT. The content-related warrant, or rationale, for the accuracy of the IA, is the process of establishing relevant knowledge and skills from representative members of the profession, and the backing in support of the warrant is the resulting analysis and blueprint. In other words, assuming the content written for an ECMT test aligned with the blueprint developed in this practice analysis study, the stated inference regarding mastery of knowledge and skills would be justified.

Taking this a step further, Kane (2006; 2013) also referred to types of inferences. The generalizability inference described in the introduction allows for observed scores on a particular set of questions or prompts to be generalized to the broader content domain. Thus, content-related evidence in the form of a practice analysis study support generalization to the broader domain of tasks performed by an ECMT and related knowledge and skills. This is supported by the results of the current study, as participants for both groups were representative of the profession across various practice settings, training background (i.e., affiliations), and demographics and the process followed industry best practices for defining the job role. Thus,

this study provides some of the content-related support necessary for generalization (knowledge, skills, and abilities –KSAs- and test content have not been developed) in the form of a documented process that led representative professionals in identifying and then validating (i.e., ratings) the most essential tasks of an ECMT from the domain of all possible tasks.

Likewise, this method via the qualifications and representation of the participants and the resulting blueprint, provide content-related backing necessary for the extrapolation of inferences from test scores to the tasks performed in classrooms, particularly representing the different settings in which teachers may work. In other words, tasks are representative of those in the various practice settings; thus, the method supports the interpretive argument for teachers in the identified settings. This does not mean competent performance on the test would represent competency on the job, only that evidence provided by this study support the assumption that test content would align with the professional role. Content-related evidence, thus, exists to support the validity argument (VA) when using results from both Panel and Survey groups. The question, then, stands of whether Panel-only results would be sufficient to support the VA.

What if there was no survey? Removing the broader survey does not change the general method. Panel experts would still rate tasks and the resulting blueprint would be the Panel results displayed in Table 4.20. Comparison of the two blueprints indicated that the resulting blueprints would be very similar, with the exception of two tasks, and even those would not produce a large difference in content of the test. Analysis of demographics indicated a strong representation of practice settings and affiliation of training programs for the Panel, and Panel representatives were carefully trained in relation to the method and goals of the rating process. The method still follows rules specified in the various standards. With regards to NCCA accreditation, Knapp, Anderson, and Wild (2015) indicated the need to provide a rationale in cases where a survey is

not included as part of the study, implying that the survey is not essential to meeting the standards. The VA may be weakened with less supporting evidence but may remain strong enough to adequately support test interpretation and use.

Kane (2006, 2013) encouraged consideration of counterarguments in examining the strength of a validity argument. The primary challenge to a practice analysis using the Panel alone as validity evidence is likely to be based on inadequate representation of the profession. In response, evidence was presented indicating that other than some imbalance in the panel across state demographic regions, the Panel more accurately represented variations in the profession due to practice settings and type of teacher training as represented by affiliation of the training program attended. The panel had a small amount of gender, racial, and ethnic diversity, similar to the survey group, possibly because this is the demographic of the profession in the state.

It could also be argued that the Panel size was too small to depend on the results alone to construct a defensible blueprint, but in this situation the population size is also small (n is estimated at approximately 185). In this small sample situation, each participant's response has greater impact on the results, thus, accuracy rests on representation (Jones, Smith, & Talley, 2006). A second consideration with regards to relying on Panel responses is that, having greater involvement in the project and more time to consider the importance of tasks, ratings are more reflective of actual practice. This has also been suggested in the literature (Kane, 1997; Raymond, 2016).

Finally, a counterargument could be made that the quality of the process used with the Panel is subjective, based only on the reporting of the testing professional who conducting the study. The panel could adequately represent the profession, but facilitation of the meeting fails to incorporate the views of all participants. Impact of dominant personalities is documented in the

literature regarding focus groups, similar in execution to a practice analysis meeting (Stewart & Shamdasani, 2014). However, an evaluation questionnaire of the in-person meeting was conducted at the end of the day in order to assess the quality of the process. Specifically, participants were asked to indicate whether they thought members had the opportunity to contribute equally. Responses indicated a strong agreement (median = 4 on a 1-4 scale) to this statement. In addition, participants indicated strong agreement that the process was defensible, providing further evidence supporting the quality of the Panel results. The instrument used to evaluate was not a validated instrument, thus future research could improve supporting evidence by using a more rigorous meeting evaluation method.

Impact on test interpretation for a specified use

Beyond potential challenges to claims based on evidence from Panel results alone, this discussion has yet to address the issue of interpretation of scores for a particular use or uses. The AG defined and Panel confirmed three possible uses of test scores. This discussion will focus on two: voluntary certification and state licensure. Recall that Kane (2006) indicated the strength of the evidence in support of a score interpretation for a particular use should be commensurate with the strength of the claim. The claim is the same, mastery of knowledge and skills required for competent performance of an entry-level ECMT, whether the purpose is voluntary certification or one component of a licensure process. What differs is the stakes involved in the use of test scores. In the first instance stakes are low to medium, with limited consequences for teachers not passing the test. There could be higher stakes for teachers if it became widely used and job opportunities were impacted by attainment of the certification.

A test required for licensure increases the stakes for teachers; thus, increasing the burden on the existing evidence. Although the broader survey is not specifically required by standards, and content-related evidence is strong in support of the interpretive argument within Kane's

framework, test use with higher stakes implies the potential for litigation. Evidence provided by this study support requirements stated by the EEOC ("Civil Rights Act of 1964," 1964), with a possible weakness due to low minority participation. However, given that the Panel group developed and validated the tasks, there may be issues for programs who wish to become accredited (National Commission for Certifying Agencies, 2016). In addition, removal of the survey could be problematic in the case of litigation as was suggested in *Gulino v. Board of Education of City School District of NY*. In this case, the court ruled that, in addition to issues with the method used to determine knowledge and skills to be tested, the survey sample was too small and lacked sufficient minority representation to provide adequate validity evidence. In addition, the court cited a related case ("*Guardians v. Civil Service Commission of the City of New York* ", 1980) indicating that the survey was a required component of a job analysis. *Guardians v. Civil Service Commission of City of New York* suggested key components of content validity evidence including identification of tasks necessary for job performance that allow for clear association of test content with the job role, as well as the administration of a survey. The current study provides evidence in support of the first, but removal of the survey portion of the study might be risky if accreditation or legal defensibility is of concern.

A second interpretation of results

Conclusions drawn, thus far, have indicated evidence in support of using results from a carefully selected panel of experts to provide adequate content-related validity evidence for stated score interpretations and uses, but may not extend to high stakes testing situations. This conclusion is based on evidence that comparison of Survey and Panel results across multiple studies indicated that opinions between groups may be close enough to warrant using Panel results only. The results that support similarities between groups also support the conclusion that the survey is doing what it is intended to do, which is verify the work of the Panel. The

expectation in a well-executed practice analysis, where essential tasks are adequately defined, is that the survey results will be similar to Panel results. This study indicates that there is little additional information provided by the survey when the panel adequately represents the professional population. Results should only vary if the Panel failed to adequately define the domain. Thus, in addition to survey results being important when stakes are high, it may also be so when evidence is weak in support of adequate professional representation on the Panel or when the process used to define tasks was weak.

Limitations of the Study

Several limitations impact the generalizability of these results. First, although this study provides a third example of research comparing Panel and Survey group ratings of task and KSA importance for competent performance in a profession, generalizing results to different conditions may not be justified. Careful consideration must be made when determining how best to conduct a practice analysis study in a given situation. Some considerations are population size, quality and representativeness of the Panel group, and intended interpretations and uses of test scores. This is particularly important in high stakes testing situations, where there may be legal concerns.

Second, this study was conducted in a single state (SC) and results of the practice analysis cannot be generalized to other states. It is possible that a comparison between Panel and Survey results in a different state or at the national level would suggest more diversity in opinions. In addition, the sample had limited involvement of urban public schools, with only small participation from experts in Richland One and no participation from the Charleston school district. This created a bias toward rural and suburban teachers and administrators. Based on feedback in the Panel meeting, which included one teacher from Richland One School District,

there appeared to be consensus among participants, but a greater representation from urban districts could have revealed differences in opinions.

Third, there was no preexisting representative database of subject matter experts to use for this study. Often practice analysis studies use a database or sample from a database of certified individuals that has been accumulated over time to survey the population or draw a sample. As there was no industry certification currently in existence, a single source did not exist. Thus, the defined population of ECMTs in the state may be an underestimation. The number of Montessori administrators was estimated based on one per school, which may not be accurate, and due to the dual role of many ECMT trainers, estimating numbers of trainers was not feasible. The demographic question regarding whether a participant was a teacher trainer, using the term “teacher educator” may have been confusing to participants; thus, the numbers of teachers answering in the affirmative are likely inflated. Although this wording was reviewed, discussed, and approved by Montessori experts, the survey participants may have interpreted teacher educator as simply a teacher.

Finally, due to very small samples there was limited utility in evaluating results for subgroups (i.e., race, gender, job role). Although Appendix P presented nearly identical blueprints for the survey group when removing certain subgroups (administrators and non-credentialed teachers and administrators), lower correlations between groups on the frequency scale may have been impacted by these or other subgroups. A future national study with larger sample sizes could investigate potential differences.

Implications for Current and Future Research

One way to address many of the limitations of this study is through additional research. In the context of current research, it was suggested that results imply justification for developing a content framework for a low to medium stakes test using Panel ratings alone. It also

demonstrated the utility of surveys in validating the work of a panel of experts, when the two groups represent the target population and the Panel has done a sufficient job of defining the domain. Future research may expand in this evidence.

Variations on the work of the combined AG/Panel could be evaluated in follow-up studies. The ability to triangulate supporting evidence is an important element of a practice analysis study that is removed when a survey is excluded from the study. As evidence does not have to be in the form of a survey, future research may focus on the impact of other sources of evidence (i.e., focus groups, interviews) in support of the work of the Panel. For example, it might be useful to use a larger AG to define initial tasks, followed by validation by a panel of experts who would review and rate these tasks. Work by the AG could be bolstered by additional methods discussed in the literature such as literature reviews, interviews, and observations. A basic review of literature conducted by a SME was included in the materials provided to the Panel, but no interviews or observations were conducted.

Related to the content of the test (i.e., defining the professional role of a Montessori teacher), future research may move the profession toward a national industry standard. Recall in Chapter 2 a summary of the current state of Montessori teacher certification. Despite an overarching accreditation for training programs, the certification landscape is disparate with credentials aligning with specific affiliates. This study was intended, in part, to inform the development of a formal, industry-wide professional certification. The first step in that development for an ECMT might be a national study to either define the role more broadly or verify the role as defined by experts in SC.

Conclusion

This study was intended to demonstrate an ability to construct adequate content-related validity evidence using a carefully selected panel of experts to determine content for a new

credentialing test. A strong reliance on survey-based research methods for determining content of credentialing tests may not always be justified, particularly when there is a small testing population. This reliance was previously challenged by two studies that examined the value of a survey beyond the work of a panel of experts: Tannenbaum and Wesley (1993) and Maurer and Tross (2000). Similar to the current study, both compared ratings from a survey of representative professionals to those from a carefully selected panel of experts, concluding that opinions between groups were similar enough to support use of the panel alone to determine the content domain of a credentialing test.

The current study sought to further investigate this issue in a situation where the testing population is very small and potentially diverse practice settings must be represented. In addition, results were presented within a theoretical framework for establishing content-related validity evidence when supporting score interpretations for purposes that may have low to high stakes. Conclusions were drawn supporting adequate evidence for low to medium stakes certification test, but require caution when selecting methods for a high stakes test that could be used for accredited programs or licensure.

In summary, research suggests that these large-scale survey-based studies may not be superior to a well-trained, focused committee or panel-based study, depending on the circumstances. Results of this study provided additional evidence in support of relying on a panel of carefully selected experts, but should be done with caution when stakes for the certified population or the public are high. In evaluating the results within the framework of Kane's (2006) argument-based approach to establishing validity, conclusions were taken a step further than previous studies, providing insight into the strength of evidence when the content

framework is not validated by a large-scale survey, and the blueprint is developed based on panel opinions alone.

APPENDIX A. SUMMARY OF PUBLISHED PRACTICE ANALYSIS STUDIES

	Author(s)/Year	Title	Method	Description
1	Althouse et al. (2009)	Confirming the validity of the General Pediatrics Certification Examinations: A practice analysis	Task Inventory Questionnaire	SME group defining/reviewing content framework followed by survey.
2	Arbet et al. (2009)	Using practice analysis to improve the certifying examinations for PAs	Task Inventory Questionnaire	SME group defining/reviewing content framework followed by survey.
3	Babcock et al (2013)	Summary of AART's 2012 practice analysis for bone densitometry	Task Inventory Questionnaire	Surveyed existing tasks from previous studies, followed by a survey
4	Barnsteiner and Wyatt (2002)	What do pediatric nurses do? Results of the Role Delineation Study in Canada and the US	Task Inventory Questionnaire	Remote SME group review of domains and tasks constructed from job logs, previous PA study, job descriptions, telephone interviews, and course syllabi, followed by a survey
5	Battenfield and Schehl (2013)	Practice analysis for human performance technologists	CIT/Task Inventory Questionnaire	Focus groups and interviews following CIT method, followed by a survey.
6	Bevans et al. (2011)	Defining clinical research nursing practice: Results of a role delineation study	Task Inventory Questionnaire	Part II of Castro et al study. Large-scale survey to validate the content framework developed in Part I.
7	Bradley et al. (2010)	A survey of North American Marriage and Family Therapy practitioners: A role delineation study	Task Inventory Questionnaire	Literature review, interviews and focus group used to establish tasks and knowledge, followed by a survey.
8	Brown et al (2012)	Oncology nurse navigator role delineation study	Task Inventory Questionnaire	SME group meeting followed by a survey.
9	Castro et al. (2011)	Validating the Clinical Research Nursing Domain of Practice	Delphi	Part I of a two part study (see Bevans et al, 2011). SME conducted literature review followed by Delphi method with three iterations.

10	Darling et al. (2009)	Professionalization of family life education: Defining the field	Task Inventory Questionnaire	SME group meeting, followed by survey
11	Willens et al. (2010)	Role delineation study of American Society for Pain Management Nursing	Task Inventory Questionnaire	SME review of existing tasks and KSAs, followed by a survey
12	Erickson et al. (2013)	Validation of holistic nursing competencies: Role-delineation study, 2012	Task Inventory Questionnaire	Literature review used to define initial tasks, followed by individual reviews by two panels and a web-based meeting with a third panel, followed by a survey
13	Garbin and Chmielewski (2013) Garbin & Chmielewski	Job analysis and role delineation: LPN/LVNs and hemodialysis technicians	Task Inventory Questionnaire	SME group meeting, followed by survey.
14	Gerrow et al. (2006)	Competencies for the beginning dental practitioner in Canada: A validity survey	Task Inventory Questionnaire	Survey using content framework developed from previous PA study
15	Gorman et al. (2010)	Nationwide acute care physical therapist practice analysis identifies knowledge, skills, and behaviors that reflect acute care practice	Task Inventory Questionnaire	SME group used literature review to define tasks and knowledge, followed by a survey
16	Hawkins-Walsh and Van Cleve (2013)	The pediatric mental health specialist: Role Delineation	Task Inventory Questionnaire	SME group defined domains and tasks, followed by a survey
17	Henman, Corrigan, Carrico, and Suh (2015)	Identifying changes in the role of the infection preventionist through the 2014 practice analysis study conducted by the Certification Board of Infection Control and Epidemiology, Inc.	Task Inventory Questionnaire	SME Groups defined task and knowledge statements, followed by a survey
18	Johanson et al. (2016)	Orthopedic physical therapy: Update to the description of specialty practice	Task Inventory Questionnaire	SME group used literature review and results of previous practice analysis to develop tasks and knowledge, followed by survey

19	Keller et al. (2016)	Thirty years of hand therapy: The 2014 practice analysis	Task Inventory Questionnaire	Tasks and KSAs developed based on telephone interviews, results reviewed by a SME panel and individuals, followed by a survey
20	Kenward (2007)	Role delineation study of nurse practitioners and clinical nurse specialists	Kane's Model-based approach/Task Inventory Questionnaire	Initial model developed based on a literature review, reviewed by SME panels, followed by a survey.
21	Koby and Melby (2013)	Certification and job task analysis (JTA) establishing validity of translator certification examinations	Task Inventory Questionnaire	Four SME focus groups defined tasks and KSAs and attitudes, followed by a survey
22	Matarese et al. (2012)	Practice analysis of nutrition support professionals	Task Inventory Questionnaire	Literature review, Interviews, SME group, survey
23	McKenzie et al. (2016)	Health Education Specialist practice analysis 2015 (HESPA 2015): Process and outcomes	Task Inventory Questionnaire	Interviews followed by SME Group review of existing content outline, reviewed remotely by a second task force, followed by a survey
24	Muckle et al. (2009)	A Report on the CCNA 2007 Professional Practice Analysis	Task Inventory Questionnaire	CCNA reviewed previous existing content framework, followed by a survey
25	Muckle et al. (2012)	Professional Practice Analysis: Determining Job Relatedness of the Certification Examination for Nurse Anesthetists	Task Inventory Questionnaire	Review of existing domains and knowledge, followed by a survey
26	Muenzen et al. (2005)	Updating the pharmacy technician certification examination: A practice analysis	Task Inventory Questionnaire	Task force reviewed previous functions and knowledge, interviewed additional SMEs, followed by a survey
27	Ortelli (2006)	Defining the professional responsibilities of academic nurse educators: The results of a national practice analysis	Task Inventory Questionnaire	SME group defined tasks, followed by a survey

28	Reuter-Rice et al. (2016)	Acute care pediatric nurse practitioner: The 2014 practice analysis	Task Inventory Questionnaire	SME group reviewed existing tasks, knowledge, and skills, followed by survey
29	Shreiner, Kolb, O'Brian, Carroll, and Lipman (2015)	National role delineation study of the board of certification for advanced diabetes management	Task Inventory Questionnaire	SME group, followed by a survey
30	Starling et al. (2014)	Job task analysis survey for the horticultural therapy profession	Task inventory Questionnaire	SME group used literature review to determined tasks and KSAs, followed by a survey
31	Tahan et al. (2015)	What case managers should know about their roles and functions	Task Inventory Questionnaire	SME Group defined domains, tasks and knowledge using literature review , followed by a survey
32	Taub et al. (2008)	A role delineation study of health educators in the United States: Methodological innovations	Task Inventory Questionnaire	Used existing competencies for a survey
33	Tsai and Kramer (2014)	Content Validity of the national board dental hygiene examination	Task Inventory Questionnaire	A joint commission determined tasks, followed by a survey
34	Universal Public Purchasing Certification Council Governance Board (2009)	Job analysis for the CPPB and CPPO Examinations	Task Inventory Questionnaire	SME review of existing content, followed by a survey
35	Vandaveer et al. (2016)	A practice analysis of coaching psychology: Toward a foundational competency model	Model-based approach/Task Inventory Questionnaire	Defined a competency model based on literature review and interviews, followed by a survey
36	Villanueva et al. (2006)	The neuroscience nursing 2005 role delineation study: Implications for certification	Task Inventory Questionnaire	Task force defined content, followed by a survey

37	Wolever et al. (2016)	Advancing a new evidence-based professional in health care: job task analysis for health and wellness coaches	Task Inventory Questionnaire	SME group defined tasks, followed by a survey
38	Yang et al	Practice Analysis : A basis for content validity for American Board of Radiology examinations in diagnostic radiology	Task Inventory Questionnaire	SME group individually defined content, followed by a survey
39	Zrebiec, J.	A national study of the Certified Diabetes Educator	Task Inventory Questionnaire	Task force created tasks using job logs from certified CDEs and existing content specifications, followed by a survey

APPENDIX B. LIST OF PARTICIPATING SCHOOLS

School	District	Est # ECMTs	Geographic Region
Abner Montessori	na	4	Midlands
Charles Towne Montessori	na	2	Lowcountry
Clemson Montessori	na	6	Upstate
Five Oaks Academy	na	6	Upstate
Lake Murray Montessori	na	5	Midlands
May River Montessori	Na	8	Lowcountry
Montessori Children’s House	na	2	Midlands
Montessori of Mt. Pleasant	na	5	Lowcountry
Montessori School of Anderson	na	9	Upstate
Montessori School of Camden	na	3	Midlands
Montessori School of Columbia	na	3	Midlands
Montessori School of Florence	na	3	Midlands
Montessori School of Greenville	na	4	Upstate
Montessori school of John's Island	na	1	Lowcountry
Montessori School of Mauldin	na	5	Upstate
Montessori School of Pawley's Island	na	5	Lowcountry
Montessori West Christian School	na	3	Upstate
Oak Grove Montessori School	na	1	Lowcountry
Sea Pines Montessori Academy	na	5	Lowcountry
Sundrops Montessori (All 3 Campus)	na	8	Lowcountry
Trinity Montessori School	na	2	Lowcountry
Barnwell Primary	Barnwell 45	2	Midlands
Boundary St. Elementary	Newberry	2	Midlands
CDC at Alfred Rush Academy	Florence One	9	Midlands
EB Morse Elementary	Laurens 55	3	Upstate
EP Todd Elementary	Spartanburg 7	3	Upstate
Fairfax Elementary	Allendale	1	Midlands
Ford Elementary	Laurens 55	1	Upstate
Gallman Elementary	Newberry	1	Midlands
Gray Court-Owings Elementary	Laurens 55	2	Upstate
Hickory Tavern Elementary	Laurens 55	1	Upstate
Hunter St. Elementary	York 1	2	Upstate
Lakeview Elementary	Greenwood 50	3	Midlands
Latta Elementary	Dillon 3	6	Midlands
Laurens Elementary	Laurens 55	3	Upstate
Lexington 4 Early Childhood Center	Lexington 4	25	Midlands
Lowcountry Montessori	State District	4	Lowcountry
McCormick Elementary	McCormick	1	Midlands
McLaurin Elementary	Florence One	9	Midlands
Newberry Elementary	Newberry	1	Midlands

School	District	Est # ECMTs	Geographic Region
North Vista Elementary	Florence One	4	Midlands
Pontiac Elementary School	Richland 2	5	Midlands
Sandy Run	Calhoun	1	Midlands
St. Matthews School	Calhoun	1	Midlands
The Children's School at Sylvia Circle	York 3	5	Upstate

APPENDIX C. DESCRIPTION OF STUDY FOR POTENTIAL PARTICIPANTS

A Certification for Montessori Educators A Proof of Concept Project

What is certification?

A certification is a test, or set of tests, designed to measure essential competencies for a particular profession or job role, such as an early childhood Montessori teacher. Independent of any specific teacher preparation program, a Montessori certification would indicate a minimum level of knowledge and skills necessary to perform competently in the classroom.

There are currently certification exams that measure a level of competency to teach, but none specific to the Montessori pedagogy or practice.

What are the benefits of certification?

There are many potential benefits of a certification specific to Montessori education for the community of teachers, students, parents, and state agencies. These include:

- An additional verification of competency independent of teacher preparation program for teachers to market their skills.
- An objective measure of competency for the hiring of Montessori teachers.
- A possible alternative to traditional teacher certification exams, such as the Praxis.
- Supports the professionalism of Montessori teachers.
- Supports the legitimacy of Montessori education and its consistent application across school settings (private, public, charter) and levels (early childhood, elementary, secondary).
- A powerful advocacy tool in seeking recognition of the Montessori teacher credential by state agencies.

What is this project?

This study is a Proof of Concept study for the design of a certification for Montessori Early Childhood educators. The initial steps indicated by testing industry standards for creating an objective certification exam will be used in South Carolina to determine viability of a study for creating such a certification nationally.

What is the study design?

There are three phases needed.

- Phase One: an Advisory team of 3-4 teachers who will work remotely with the researcher to collect information on the role of the Montessori teacher.
- Phase Two: A one day, face to face, meeting with 8-12 teachers and teacher educators to identify the knowledge and tasks necessary for competent performance in the Montessori classroom.
- Phase Three: A survey completed by the accessible population of Montessori Early Childhood teachers in South Carolina who will be asked to rate each item on its importance for the entry level Montessori teacher.

Do you want to benefit Montessori education?

Volunteer!

Phase One:

- **The Work:** Attend 2-3 1-hour web-based meetings to determine initial information about Early Childhood Montessori teachers in South Carolina and assist in drafting some initial tasks statements related to the work these teachers do in the classroom. These advisors will also be asked to complete the brief survey that will be administered in Phase three of the study.

- **The Benefits:** Involvement in the development of a professional certification for Montessori teachers; network with colleagues; great leadership opportunity.

Phase Two:

- **The Work:** attend a one day meeting on January 9th in the Columbia area to participate in identifying the competencies, knowledge and skills needed by the entry level Montessori Early Childhood teacher. Hotel accommodations and meals will be provided.
- **The Follow up:** complete a brief survey and attend a 1 hour follow up web meeting to review study results.
- **Eligibility:** Hold an Early Childhood Montessori credential from a MACTE accredited Teacher Preparation Program; 3-8 years of Montessori teaching experience in SC.
- **The Benefits:** Involvement in the development of a professional certification for Montessori teachers; network with colleagues; great leadership opportunity.

Phase Three:

- **The Work:** Complete an online survey (approximately 20 minutes)
- **The Benefits:** Contribute to the development of a certification for Montessori educators.

If you are interesting in participating in this study please contact Diane Talley.

Diane Talley
Doctoral Candidate
University of North Carolina, Chapel Hill
919-410-7792
dtalley@live.unc.edu

<https://www.linkedin.com/in/dianetalley>

Text for email invitation to the Target audience: Representatives of Montessori associations, teacher education programs, and schools

Dear Montessorian,

You may already have heard about a proposed research study to establish the foundation for a Montessori teacher certification program. I would like to share the details of this study and the benefits it will have for Montessori education and its community. I now have approval to move forward with this research and am beginning to recruit teachers and teacher educators in South Carolina to participate in the study.

An analysis of a professional practice such as the Montessori teaching profession is a well-established method for determining the content of a certification test. This endeavor, however, requires a partnership between professional experts, which we call subject matter experts, and someone who understands test development. The professional experts who participate in such a study are invaluable in generating quality, usable results.

I would like the opportunity to provide more details regarding this study and what expertise I am seeking in conducting this work. Please let me know if you would like to schedule a time to speak.

Best regards,

Diane Talley

Doctoral Candidate

University of North Carolina, Chapel Hill

dtalley@live.unc.edu

APPENDIX D. CONSENT FORM FOR ADVISORY GROUP

University of North Carolina-Chapel Hill

Consent to Participate in a Research Study Phase I: Practice Analysis Subject Matter

Expert Advisory Committee

IRB Study # 15-2595

Consent Form Version Date: 9/29/2015

Title of Study: Early Childhood Montessori Teacher Practice Analysis

Principal Investigator: Diane Talley

UNC-Chapel Hill Department: School of Education Deans Office

Co-Investigators:

Funding Source:

Study Contact: dtalley@live.unc.edu

What are some general things you should know about research studies?

You are being asked to take part in a research study. To join the study is voluntary. You may refuse to join, or you may withdraw your consent to be in the study, for any reason, without penalty.

Research studies are designed to obtain new knowledge. This new information may help people in the future. You may not receive any direct benefit from being in the research study. There also may be risks to being in research studies.

Details about this study are discussed below. It is important that you understand this information so that you can make an informed choice about being in this research study. You will be given a copy of this consent form. You should ask the researchers named above, or staff members who may assist them, any questions you have about this study at any time.

What is the purpose of this study?

The purpose of this research study is to define the critical tasks of an Early Childhood Montessori Teacher necessary for competent performance in a Montessori classroom. The primary goal of collecting such information is to create a content domain and blueprint from which a certification test may be developed at some point in the future.

How many people will take part in this study?

If you decide to be in this study, you will be one of approximately 16 people in the first two phases of the research study, which is being conducted in South Carolina. Participants in this part of the research have expertise in the role of an Early Childhood Montessori Teacher and were carefully selected based that expertise and level of experience in the field.

How long will your part in this study last?

Your participation in this focus group will last approximately one hour.

What will happen if you take part in the study?

The group will be asked to discuss how the mitigation process works in the various communities, and how mitigation participation might be improved. No questions will be directed to you individually, but instead will be posed to the group. You may choose to respond or not respond at any point during the discussion. The focus group discussion will be audiotaped so we can capture comments in a transcript for analysis.

What are the possible benefits from being in this study?

Research is designed to benefit society by gaining new knowledge. You may not benefit personally from being in this research study.

What are the possible risks or discomforts involved from being in this study?

We do not anticipate any risks or discomfort to you from being in this study. Even though we will emphasize to all participants that comments made during the focus group session should be kept confidential, it is possible that participants may repeat comments outside of the group at some time in the future. Therefore, we encourage you to be as honest and open as you can, but remain aware of our limits in protecting confidentiality.

How will information about you be protected?

Every effort will be taken to protect your identity as a participant in this study. You will not be identified in any report or publication of this study or its results. Your name will not appear on any transcripts; instead, you will be given a code number. The list which matches names and

code numbers will be kept in a locked file cabinet. After the focus group tape has been transcribed, the tape will be destroyed, and the list of names and numbers will also be destroyed.

Will you receive anything for being in this study?

You will not receive anything for taking part in this study.

Will it cost you anything to be in this study?

There will be no costs for being in the study

What if you are a UNC employee?

Taking part in this research is not a part of your University duties, and refusing to participate will not affect your job. You will not be offered or receive any special job-related consideration if you take part in this research.

What if you have questions about this study?

You have the right to ask, and have answered, any questions you may have about this research. If you have questions, or concerns, you should contact the researchers listed on the first page of this form.

What if you have questions about your rights as a research participant?

All research on human volunteers is reviewed by a committee that works to protect your rights and welfare. If you have questions or concerns about your rights as a research subject you may

APPENDIX E. PHASE I. ESTABLISHING FOUNDATIONAL INFORMATION FOR THE PRACTICE ANALYSIS PANEL

Goal of Phase I

The goal of Phase I is to establish foundational information and materials to be used in Phase II with the practice analysis panel. This information includes, but is not limited to the following:

- Existing standards for early childhood teachers (training, assessment, certification, job requirements/descriptions)
- Job descriptions for early childhood Montessori teachers, particularly in South Carolina (may come from schools, state agencies, Montessori organizations)
- References related to the role of a teacher in a Montessori classroom

Materials to develop

1. Comparisons between certification requirements of individual training programs in South Carolina
2. Draft of a statement describing the target audience (job role, education, level of experience, practice settings)
3. Proposed eligibility requirements
4. Proposed test format and justification
5. Examples of task statements
6. Sample survey with examples of tasks, instructions, and measurement scales
7. Literature review on the role of an early childhood Montessori teacher

Description of work process

The role of this advisory committee is to support the efforts of the psychometrician in preparing information and materials for the larger practice analysis panel. Meetings will be

conducted remotely. There are 3 primary meetings planned for this group in addition to email and phone communications:

1. An instructional meeting to inform the committee of basic concepts of practice analysis methodology, the role of the committee and the purpose of the practice analysis panel, and the scope of work for Phase I of the project. – 2 hours
2. Brainstorming session to discuss target audience, eligibility requirements, format of the test, and comparison of these with the Praxis PLT for early childhood. Instructions for how to construct task statements to align with the target audience. – 2 hours
3. Review draft task statements and wrap up. – 2 hours

Qualifications and time commitments of advisory committee

The ideal advisory committee members meet the following criteria:

- 3-8 years of experience as an early childhood Montessori teacher (At least 2 years teaching in SC)
- Currently teaching in a public or private Montessori classroom in South Carolina
- Credentialed by a MACTE accredited training program
- Be available for 3 2-hour, web-based meetings and approximately 6 hours of work outside of meetings (approximate time commitment is 12 hours)

Advisory Group Meeting 1 Agenda

- Introductions (2-3 minutes)
- Review purpose of the advisory group (3-5 minutes)
- Review key concepts (15 minutes)
- Purpose and format of the test (5 minutes)
- Group activity: Describe a competent early childhood Montessori teacher (25 minutes)
 - We're not describing what they do yet (tasks), just talking about education, experience, other training, certifications/certificates, affiliates. How do we describe this person?
- Eligibility requirements: What will teachers have to do to be able to sit for the test based on the description?
- Quick glimpse at the next task (5 minutes)
- Introductions (2-3 minutes)
- Review purpose of the advisory group (3-5 minutes)
- Review key concepts (15 minutes)
- Purpose and format of the test (5 minutes)
- Group activity: Describe a competent early childhood Montessori teacher (25 minutes)
 - We're not describing what they do yet (tasks), just talking about education, experience, other training, certifications/certificates, affiliates. How do we describe this person?
- Eligibility requirements: What will teachers have to do to be able to sit for the test based on the description?
- Quick glimpse at the next task (5 minutes)

Practice Analysis Advisory Committee: Meeting 1

Tuesday, November 10, 2015

In attendance:

Diane Talley – Researcher

Christine Lowry- Montessori Now, project advisor

Regina Colins – EC Teacher, Latta Elementary School

Donna Brown – EC Teacher Early Childhood Education Center, Lexington 4 school district

Meek Duran – EC Teacher, Montessori school of Camden

Karen Toliver - EC Teacher, Montessori school of Camden

Oscar Avila - EC Teacher, Montessori school of Camden

John Moncure – Head of School, Montessori school of Camden

Note: ECMT will be used to denote Early childhood Montessori teacher

Purpose of the meeting: To clarify the goals and methods of the study, define study-related terminology, define the target testing population (a just qualified ECMT), and proposed eligibility requirements for an ECMT test.

Purpose of the certification test: The purpose of the test is to provide an independent

verification of qualifications for an ECMT. The possible uses of this test are:

- A voluntary credential for all teachers that may be used to demonstrate competency and differentiate themselves professionally, as well as for hiring managers to use as a verification of competency independent of the teacher preparation programs and affiliate-specific credentials.

- A mandatory certification used by state agencies to assess competency of an entry-level ECMT seeking licensure to work in ECM public education programs.
- Provide cohort-level feedback for teacher preparation programs in evaluating the competency of their graduates.

Definition of a just qualified ECMT: A just qualified ECMT is an entry-level teacher who has graduated from a MACTE accredited early childhood preparation program, which requires 200 hours of academic work and 400 hours of practicum work. Teachers are prepared at a minimum in areas of practical life, sensorial, mathematics, language, science, physical geography, cultural studies, physical education, music, and arts education.

The ECMT has experience as a supervised teacher during the practicum portion of their education, but is not expected to have unsupervised experience in the ECM classroom. A just qualified ECMT will have knowledge of the basic materials in a Montessori classroom. The fundamental materials necessary for an entry-level teacher to be able to use are:

Practical Life:

Hand washing	Object Scrubbing	Sweeping
Dry pouring	Wet pouring	Spooning
Tonging	Sorting	Food preparation
Sewing	Dressing frames	Cutting and gluing

Sensorial:

Knobbed cylinders	Tower of cubes	Broad Prisms
Length Rods	Color Box 1-3	Constructive Triangle boxes
Knob less cylinders	Touch Boards	Sound cylinders

Geometric solids	Geometric cabinet	Stereognostic exercises
Mystery bag	Smelling exercises	Binomial Cube
Trinomial Cube		

Language:

Metal insets	Sandpaper letters	Moveable alphabet
Vocabulary activities	Reading Readiness	Objects/moveable alphabet
Pictures/moveable alphabet	Phonogram exercises	Phonetic Reading
Puzzle Words		

Math:

Number rods	Sandpaper numerals	Spindle Boxes
Cards and counters	Golden Bead materials	Stamp Game
Bead Frames	Ten Board	Teen Board
Hundred Board	Addition Strip Board	Addition Charts
Subtraction Strip Board	Subtraction Charts	Multiplication Board
Multiplication Charts	Division Board	

Cultural:

Globes	Land Forms	Maps
Cultural Geography	History	Plant puzzles
Botany 3 part cards	Animal puzzles	Zoology 3 part cards
Art materials	Library area	Music area

Eligibility requirements: In order to qualify to take the ECMT, teachers must fulfill the following requirements:

- Certification from a MACTE accredited teacher preparation program
- Minimum 400 hours of supervised teaching
- Bachelor's degree is recommended, but not required

Next Step: Create a content structure for the test

The following is an example of a content outline with domains and tasks from a Certified Specialty Pharmacist practice analysis study. This is an excerpt from a much larger list, but is intended to give you an idea of what we are trying to create for an ECMT. The roman numerals represent high-level categories (domains) and the numbered sections below each domain are tasks.

Intake
Screen patient demographic and clinical information for suitability for specialty pharmacy services
Verify completeness of clinical information required (e.g. diagnosis code) to ensure that prescription is appropriate
3. Obtain patient attestation as required by regulations (e.g. HIPAA, REMS)
Clinical Management
Establish, document, and update patient clinical record
Assess clinical data for relevant characteristics that affect effectiveness of medication and associated risks (e.g. comorbidities, contraindications, duplicative therapy, allergies age, weight)
Follow up with patients individually to assess treatment progress and quality of life
Fulfillment
Verify that medication is available

Refer medication referral to another provider if unable to distribute medication
Dispense the medication prescription
Outcomes
Determine clinical, patient-reported, operational, and financial data to be collected based on the parameters of disease state and medication, and how data will be obtained from internal and external sources

The next step in this process is to define the content domains. Here is an example of what content domains might be for an ECMT content outline.

I. Montessori Child Development and Philosophy

II. Learning and Teaching

III. Curriculum

IV. Classroom Management

V. Families and Communities

VI. Professionalism

Advisory Group Meeting 2 Agenda

- Where we are in the project – 5 min
- Overview of content outline- 5 min
- Need 2-3 good examples for the January 9 meeting
- Task exercise- 10min
- Domains – 20 min

- Anatomy of a task and plan for developing examples – 10 min
- Review of basic materials list – 10 min

Practice Analysis Advisory Committee: Meeting 2

Tuesday, December 8, 2015

In attendance:

Diane Talley – Researcher

Christine Lowry- Montessori Now, project advisor

Regina Colins – EC Teacher, Latta Elementary School

Meek Duran – EC Teacher, Montessori school of Camden

Karen Toliver - EC Teacher, Montessori school of Camden

Oscar Avila - EC Teacher, Montessori school of Camden

Not present:

John Moncure – Head of School, Montessori school of Camden

Donna Brown – EC Teacher Early Childhood Education Center, Lexington 4 school district:

Donna provided comments by email in advance of the meeting.

Note: ECMT will be used to denote Early childhood Montessori teacher

Purpose of the meeting: To establish content domains and 2-3 sample tasks for use in the panel meeting on January 8.

Review from last meeting:

Purpose of the certification test: The purpose of the test is to provide an independent verification of qualifications for an ECMT. The possible uses of this test are:

- A voluntary credential for all teachers that may be used to demonstrate competency and differentiate themselves professionally, as well as for hiring managers to use as a verification of competency independent of the teacher preparation programs and affiliate-specific credentials.
- A mandatory certification used by state agencies to assess competency of an entry-level ECMT seeking licensure to work in ECM public education programs.
- Provide cohort-level feedback for teacher preparation programs in evaluating the competency of their graduates.

Definition of a just qualified ECMT: A just qualified ECMT is an entry-level teacher who has graduated from a MACTE accredited early childhood preparation program, which requires 200 hours of academic work and 400 hours of practicum work. Teachers are prepared at a minimum in areas of practical life, sensorial, mathematics, language, science, physical geography, cultural studies, physical education, music, and arts education.

The ECMT has experience as a supervised teacher during the practicum portion of their education, but is not expected to have unsupervised experience in the ECM classroom. A just qualified ECMT will have knowledge of the basic materials in a Montessori classroom.

The list of materials needs to be reviewed by the group with Donna's comments. This will be posted in Google Docs for group review prior to the January 9 meeting.

Eligibility requirements: In order to qualify to take the ECMT, teachers must fulfill the following requirements:

- Certification from a MACTE accredited teacher preparation program
- Minimum 400 hours of supervised teaching
- Bachelor's degree is recommended, but not required

Draft Content Domains: There were discussed and agreed upon by the group, but should be reviewed again by the advisory group and then by the panel. **The domains are provided in the shared Google doc for the advisory group’s final review. We also need to elaborate on these descriptions.**

Example of what we are trying to achieve:

Domain	<u>Assessment</u>: Evaluating students in order to meet individual learning needs. This includes formative and summative evaluations.
Task	Evaluate mastery level of individual students using formative assessment techniques in order to provide differentiated instruction.
Knowledge of	Formative assessment techniques Levels of mastery for individual works
Skills in	Applying formative assessment techniques to evaluate mastery. Determining specific instructional needs of students based on observations.

- The goal of this project is to develop Domains and Tasks. I’ve included Knowledge and Skills to demonstrate the distinction between tasks (what you do) and knowledge and skills (what you need to know in order to “do” that task).

Content Domains

Individualized Instruction – this would include tasks related to teaching a lesson, make sure there is a controlled error, assessment, guiding learning

The prepared environment – arranging work, make it beautiful, set the tone of the environment

Classroom Management – Peace curriculum, behavior management, setting boundaries/ground rules, grace and courtesy

Assessment – formative assessment (observations, documentation), goal setting (may not be across all schools), summative assessment (report cards/narratives), example: Kauffman test, PALS (2 times per year), DRA2

Communications - Reporting Parent/teacher conferences, other teachers, administrators, communications with school community, parent education (may not be for a new teacher), provide tours to prospective parents, manage parent observations

Tasks

To establish tasks, we began by listing in simple form things a new ECMT would need to be competent doing in the classroom. The follow is that list:

1. Position works on shelves
2. Select developmentally appropriate materials
3. Set the tone for the classroom
4. Give a lesson
5. Evaluate the child
6. Make sure there is a controlled error
7. Individualize instruction

Next, we talked about how to write formal task statements:

Example: Evaluate mastery level of individual students using formative assessment techniques in order to provide differentiated instruction.

- **Verb** - Evaluate
- **Direct object**– evaluating what?
- **Using what tool(s)** – formative assessment techniques
- **For what purpose** – provide differentiated instruction

We wrote an example as a group. This example is in the Google doc online. Some things to keep in mind when you are constructing tasks, knowledge and skills:

- Tasks must be something you can measure objectively and in a standardized way.
- Avoid soft skills such as timeliness, organized, pleasant to work with. These must be evaluated outside of this type of test.
- Keep in mind that when test items are written, each item measures one general concept which aligns with a task and knowledge or skill. The task statement, thus, should focus on one discrete task, such as giving a lesson. The same goes for knowledge and skills. If you want a teacher to have the knowledge of what the lesson is and how to use it, these are two separate things. The first is knowledge, the second is skill in how to use that information. Some more complex items might combine knowledge and skills, but in general each statement should be focused on one or the other.
- Tasks should be a single statement, not broken down into multiple sentences.
- When you consider how you perform a task, do you perform it in different ways for different groups? Would you teach a lesson differently for individuals and groups? If so these should be two separate tasks. Think about the knowledge and skills you need to

perform that task. Is there something different you need to know pedagogically when interacting with an individual versus a group?

- Use action verbs such as teach, guide, assess, set up, arrange, instruct, demonstrate rather than less clear verbs such as assist, support, consider, understand. Consider what *assist* means; what are you doing when you are *assisting*?
- Use caution when using the word prepare. We need to rethink the task that states “prepare the environment.” This is very broad and includes multiple sub-tasks. What do you actually do when you prepare; do you set up, organize, select, arrange? We only need one of these as an example at this point.

The task to complete prior to January 8 is to review and revise the domains and the sample tasks, knowledge and skills. I have put the example that I wrote (just as an example, not to be used as a final task) as well as the work we did on Tuesday in a Google doc file. Christine has written a draft of the other two sections so all you have to do is review and give feedback. If you feel compelled to write an additional example please do! Here’s how you access and review:

- To access, go to the email that I sent you and click on the name of the Google document. It should take you directly to it.
- Once you are in the document, look at the upper right hand side for a menu that says *Editing*. Select *suggestion*. This works more or less like track changes.
- If you want to make a comment, right click on the place where you want to make a comment and a box will pop up. You can also reply to other people’s comments by clicking on their comment box and hit reply.

There is one last task, mentioned earlier. Donna Brown made some comments on the materials list and I need your feedback please! I will put this list at the bottom of the Google Doc we are working in.

Review of literature written by Montessori on the Role of the Montessori Teacher

The Prepared Environment:

- Prepare a dynamic, interesting environment of activities
- Prepare environment that allows child to master activities and develop concentration; development of auto-education
- Provide order; a place for every object; sequence of lessons
- Environment should provide very few contrasting stimuli
- Teacher is the keeper and custodian of the environment; keep materials in order; care for the environment

(Montessori, 1916; Montessori, 1936; Montessori, 1948; Montessori, 1949; Montessori, 2008)

Assessment and Observation:

- Observe concentration, child's interest and length of interest, attention and repetition
- Assess readiness for lessons; ability to generalize
- A lesson should be followed with observation to assess development of the concept
- Know when to observe and not interrupt the child; recognize deep concentration
- Carefully observe all children to offer help when needed and know when it is not needed
- Observe child's physical, cognitive, and social-emotional development
- "The way in which we observe a child is extremely important. It is not sufficient to have a merely theoretical knowledge of education." (Montessori. Secret of Childhood. P. 49)

(Montessori, 1916; Montessori, 1936; Montessori, 1948; Montessori, 1949; Montessori, 2008)

Interaction: with children

- Teacher should refrain from interfering, interrupting, praising, or correcting child who is concentrating in a manner that furthers his own development
- Teacher should allow the child to become independent of her assistance
- Teacher should listen and reciprocate and go to child who requests assistance
- Manage behavior which is disorderly and chaotic through redirection

(Montessori, 1912; Montessori, 1936; Montessori, 1948; Montessori, 1949; Montessori, 2008)

Behavior/Classroom Management:

- Importance of ground rules and grace and courtesy lessons; restrict use of a material prior to a lesson
- Close observation to notice difference between impulsive behavior vs. that which furthers child's development
- Awareness/responsiveness to lack of coordinated movement with negative behaviors, inability to focus, poor impulse control- provide management techniques that correct these behaviors
- Management techniques should be based on judgment and intelligence; knowledge of developmentally appropriate behavior; understand the causes of disordered behavior
- Use appropriate authority as needed; important to stop undisciplined behavior
- Help child to end an activity when it is no longer being used in a way that furthers development
- Encourage independence; do not give unnecessary help; interact with respect that does not impose adult will

- Refrain from interfering in children's social interactions; allow children to solve own problems

(Montessori, 1912; Montessori, 1916; Montessori, 1936; Montessori, 1948; Montessori, 1949; Montessori, 2008)

Individualized Instruction:

- Have a good knowledge of the function of each material; the sequence of the curriculum
- Use the three period lesson format
- Teach exact names and vocabulary
- Have a good knowledge of what lesson or activity is developmentally appropriate
- Know when to give a specific lesson to an individual child
- Guide and draw children to meaningful work; be the link between the child and the material
- Give individual lessons- explain use of material; teach use of objects; isolate object and demonstrate presentation; use minimal language, focus on material not teacher; do not correct errors; simple to complex- one concept at a time; concrete to abstract; left to right; control of error
- Provide individualized support
- Teach positive use of material and practical tasks to further the child's development

(Montessori, 1916; Montessori, 1936; Montessori, 1948; Montessori 1949; Montessori, 2008)

General Knowledge and Teaching Techniques and Teacher Characteristics:

- Practical experience is necessary
- Good understanding of child development; Montessori Planes of Development; sensitive periods

- Intentional teaching based on assessment of developmental levels of individual child
- Teacher should be calm, firm, and patient; be a role model in appearance and movement; know when to observe and when to intervene; do not disturb a child who is concentrating; allow child to make choices independent of the teacher
- The Montessori teacher should strive to develop tranquility, patience, humility, open minded curiosity; have a scientific attitude; master and control her own will
- The Montessori teacher should regularly engage in self-reflection

(Montessori, 1916; Montessori, 1936; Montessori, 1948; Montessori, 1949; Montessori, 2008)

Montessori, M. (1912/1964) *The Montessori Method*. New York, NY: Schocken Books.

Montessori, M. (1916/2004) *The Advanced Montessori Method-1*. Oxford, England: Clio Press

Montessori, M. (1936/1982) *The Secret of Childhood*. New York, NY: Ballantine Books

Montessori, M. (1948/1986) *The Discovery of the Child*. New York, NY: Random House

Montessori, M. (1949/1995) *The Absorbent Mind*. New York, NY: Holt and Company, LLC

Montessori, M. (2008) *The Child, Society and the World: Unpublished speeches and writings*.

Amsterdam, The Netherlands: Montessori-Pierson Publishing Company

APPENDIX F. PANEL MEMBER CONSENT FORM

University of North Carolina-Chapel Hill

**Consent to Participate in a Research Study Phase II: Practice Analysis Subject Matter
Expert Panel**

IRB Study # 15-2595

Consent Form Version Date: 9/29/2015

Title of Study: Early Childhood Montessori Teacher Practice Analysis

Principal Investigator: Diane Talley

UNC-Chapel Hill Department: School of Education Deans Office

Co-Investigators:

Funding Source:

Study Contact: dtalley@live.unc.edu

What are some general things you should know about research studies?

You are being asked to take part in a research study. To join the study is voluntary. You may refuse to join, or you may withdraw your consent to be in the study, for any reason, without penalty.

Research studies are designed to obtain new knowledge. This new information may help people in the future. You may not receive any direct benefit from being in the research study. There also may be risks to being in research studies.

Details about this study are discussed below. It is important that you understand this information so that you can make an informed choice about being in this research study. You will be given a copy of this consent form. You should ask the researchers named above, or staff members who may assist them, any questions you have about this study at any time.

What is the purpose of this study?

The purpose of this research study is to define the critical tasks of an Early Childhood Montessori Teacher (ECMT) necessary for competent performance in a Montessori classroom. The primary goal of collecting such information is to create a content domain and blueprint from which a certification test may be developed at some point in the future.

How many people will take part in this study?

If you decide to be in this study, you will be one of approximately 12 participants in this phase of the research study, which is being conducted with teachers and teacher educators in South Carolina. Participants in this part of the research have expertise in the role of an ECMT and were carefully selected based that expertise and level of experience in the field.

How long will your part in this study last?

Your participation in this focus group will last approximately 3 months.

What will happen if you take part in the study?

You will be asked to participate in 2 hour web-based meeting to prepare for the in person panel discussion, a 1-day in person meeting in central South Carolina, and some potential subsequent hours remotely to complete the draft content domain. After the content has been surveyed, there may be an additional 2 hour conference call to discuss the results and make alterations to the final content of the study.

What are the possible benefits from being in this study?

Research is designed to benefit society by gaining new knowledge. This research is intended to benefit Montessori education and the early childhood Montessori teaching profession by establishing a clearly defined description of the professional domain and provide the foundation for a potential certification test. You may not benefit personally from being in this research study, although participating as a subject matter expert in practice analysis study is, in many industries, considered a prestigious role and many individuals include such participation on resumes.

What are the possible risks or discomforts involved from being in this study?

We do not anticipate any risks or discomfort to you from being in this study. Even though we will emphasize to all participants that comments made during group discussions should be kept confidential, it is possible that participants may repeat comments outside of the group at some time in the future. Therefore, we encourage you to be as honest and open as you can, but remain aware of our limits in protecting confidentiality.

How will information about you be protected?

Every effort will be taken to protect your identity as a participant in this study. A record of participants in this phase of the research is kept along with the qualifications (i.e., education and years of experience) and basic demographic information (i.e., city and school affiliation) as part of the research results. However, you will not be identified in any publications of this study.

Will you receive anything for being in this study?

You will not receive anything for taking part in this study.

Will it cost you anything to be in this study?

There will be no costs for being in the study.

What if you have questions about this study?

You have the right to ask, and have answered, any questions you may have about this research. If you have questions, or concerns, you should contact the researcher listed on the first page of this form.

What if you have questions about your rights as a research participant?

All research on human volunteers is reviewed by a committee that works to protect your rights and welfare. If you have questions or concerns about your rights as a research subject you may contact, anonymously if you wish, the Institutional Review Board at 919-966-3113 or by email to IRB_subjects@unc.edu.

APPENDIX G. MEETING WORKBOOK

Early Childhood Montessori Teacher Certification Practice Analysis Study

Saturday, January 9, 2016

Camden, South Carolina

Meeting facilitated by Diane Talley

Panel Participant Name _____

Meeting Agenda

- 8:30-9:00** Continental Breakfast and Introductions
- 9:00-9:20** Goals, ground rules, and what is a practice analysis
- 9:20-10:00** Large group discussion of test purpose, target audience, and eligibility
- 10:00-10:20** Domains, tasks, knowledge, skills, and items
- 10:20-12:00** Small group task activity
- 12:00-12:30** Break for lunch
- 12:30-1:30** Large group review of tasks and domains
- 1:30-4:00** Small group task revisions and large group review
- 4:00** Wrap up and next steps

Practice Analysis Advisory Committee

The purpose of this committee is to assist the researcher in preparing initial information and materials for the in-person panel meeting.

Christine Lowry

Karen Toliver

Meek Duren

John Moncure

Donna Brown

Regina Colins

Practice Analysis Panel

The primary purpose of the panel is to identify the critical tasks of the ECMT.

Oscar Avila

Regina Colin

Bethany Counts

Barbara Ervin

Kelly Gant

Phyllis Hyatt

Candace Kirby

Laura Lopresto-MCKeown

John Moncure

Carla Moore

Teresa Noble

Melissa Redd

Missy Sawyer

Rebecca Williams

Terminology

Assessment-based certificate: A credential that requires candidates complete some form of training followed by an assessment, often with a designated period of time or practicum required prior to the assessment.

Certification: A process for demonstrating competency in a profession or job role that requires test candidates to meet certain educational and experiential (eligibility) requirements, and then

pass a test or series of tests in order to achieve certification. Strictly speaking, a certification test is independent of training or education.

Domain: A high level category in the classification of tasks performed by individuals in a profession or job role. Its function is primarily to create an organized structure for the content domain of a test, but can, in some circumstances, be used in defining sections for subscores on tests.

ECMT: Acronym used in this study to indicate an Early Childhood Montessori Teacher.

Just qualified teacher (JQT): A JQT is a teacher who has demonstrated that he or she possesses the minimum knowledge and skills to competently perform in a job role. Certification tests are designed to measure candidates against a standard that represents this minimum level of competency. In a practice analysis study, we define what a JQT is and the essential tasks, knowledge, and skills they must possess in order to demonstrate minimum competency.

Knowledge, skills and abilities (KSAs): KSAs are the knowledge, skills, and abilities necessary for competent performance in a profession or job. Test content (items, tasks, prompts, etc) will align with these KSAs (see figure on page 8).

Licensure: A process for obtaining a credential in a government regulated profession that typically includes education/training as well as passing a test or series of tests in order to demonstrate minimum competency.

Practice analysis study: A systematic method for establishing a content framework, made up of the critical competencies of the profession, for a credentialing test. A practice analysis provides the foundation for a testing program, key evidence of validity, and legal defensibility (See page 5 for additional information).

Target population: The audience for whom a certification is intended. A target population should be defined in terms of educational background, experience, geographical representation, and practice settings.

Task: That which a professional does on the job, such as give a lesson, assess a student's comprehension, or communicate classroom behavior expectations to students. A task is distinguished from the knowledge, skills, and abilities necessary to competently execute a task. In order to give a lesson, a teacher must have knowledge of content and child development, and skills in applying appropriate instructional methods.

Validity: The extent to which the evidence accumulated in the process of planning, developing, and administering a test support the interpretation of test scores for a particular use (AERA et al., 2014). In the case of a teacher certification, a passing score may indicate that an entry-level teacher has the minimum knowledge and skills to teach independently in the classroom. Thus, we may say that the scores are valid for use in determining a teacher's competency for entry into the profession. Scores may not be valid for determining a teacher's advanced skills in the classroom since the test was not developed to measure those advanced skills.

Practice Analysis Study

A practice analysis study is a systematic method for establishing a content framework, made up of the critical competencies of the profession, for a credentialing test. There are many different methods for accomplishing this, the most common being a three phase study.

- I. Planning: Gathering initial information regarding the job role. This also includes recruiting subject matter experts (SMEs) to participate in Phases II and III of the study. The planning work for the current study was completed by the research with assistance

from Christine Lowry of Montessori Now and a small group of South Carolina SMEs (the advisory committee).

- II. In-person meeting: A group of 8-12 SMEs, representative of the profession, meet for 1-3 days to define essential tasks and KSAs necessary for competent performance on the job. The current project, since the meeting is only one day, will focus on defining the tasks only.
- III. Survey: A questionnaire administered to a representative sample of professionals in order to validate the content defined by the panel. An identical, but separate, questionnaire will also be administered to the SME panel.

ECMT test purpose, target audience, and eligibility requirements

- ❖ **Large group activity.** These definitions were drafted by the advisory committee as a starting point for the panel. We will review and revise the purpose, definition of target audience, and eligibility requirements in this activity.

Purpose of the certification test: The purpose of the test is to provide an independent verification of qualifications for an ECMT. The possible uses of this test are:

- A voluntary credential for all teachers that may be used to demonstrate competency and differentiate themselves professionally, as well as for hiring managers to use as a verification of competency independent of the teacher preparation programs and affiliate-specific credentials.
- A mandatory certification used by state agencies to assess competency of an entry-level ECMT seeking licensure to work in ECM public education programs.

- Provide cohort-level feedback for teacher preparation programs in evaluating the competency of their graduates.

Definition of a just qualified ECMT: A just qualified ECMT is an entry-level teacher who has graduated from a MACTE accredited early childhood preparation program, which requires 200 hours of academic work and 400 hours of practicum work. Teachers are prepared at a minimum in areas of practical life, sensorial, mathematics, language, science, physical geography, cultural studies, physical education, music, and arts education.

The ECMT has experience as a supervised teacher during the practicum portion of their education, but is not expected to have unsupervised experience in the ECM classroom. A *just qualified* ECMT will have knowledge of the basic materials (see list on page 8) in a Montessori classroom.

Eligibility requirements: In order to qualify to take the ECMT, teachers must fulfill the following requirements:

- Certification from a MACTE accredited teacher preparation program
- Minimum 400 hours of supervised teaching
- Bachelor’s degree is recommended, but not required

Fundamental materials necessary for an entry-level teacher to be able to use are:

Practical Life:

Hand washing	Object Scrubbing	Sweeping
Dry pouring	Wet pouring	Spooning
Tonging	Sorting	Food preparation
Sewing	Dressing frames	Cutting and gluing

Sensorial:

Knobbed cylinders/Cylinder blocks	Tower of cubes/Pink tower	Broad Prisms/Brown stairs
Length Rods/Red rods	Color Box 1-3	Constructive Triangle boxes
Knob less cylinders	Touch Boards	Sound cylinders
Geometric solids	Geometric cabinet	Tasting bottles
Mystery bag	Smelling bottles	Binomial Cube
Trinomial Cube		

Language:

Metal insets	Sandpaper letters	Moveable alphabet
Objects/moveable alphabet	Pictures/moveable alphabet	Grammar box

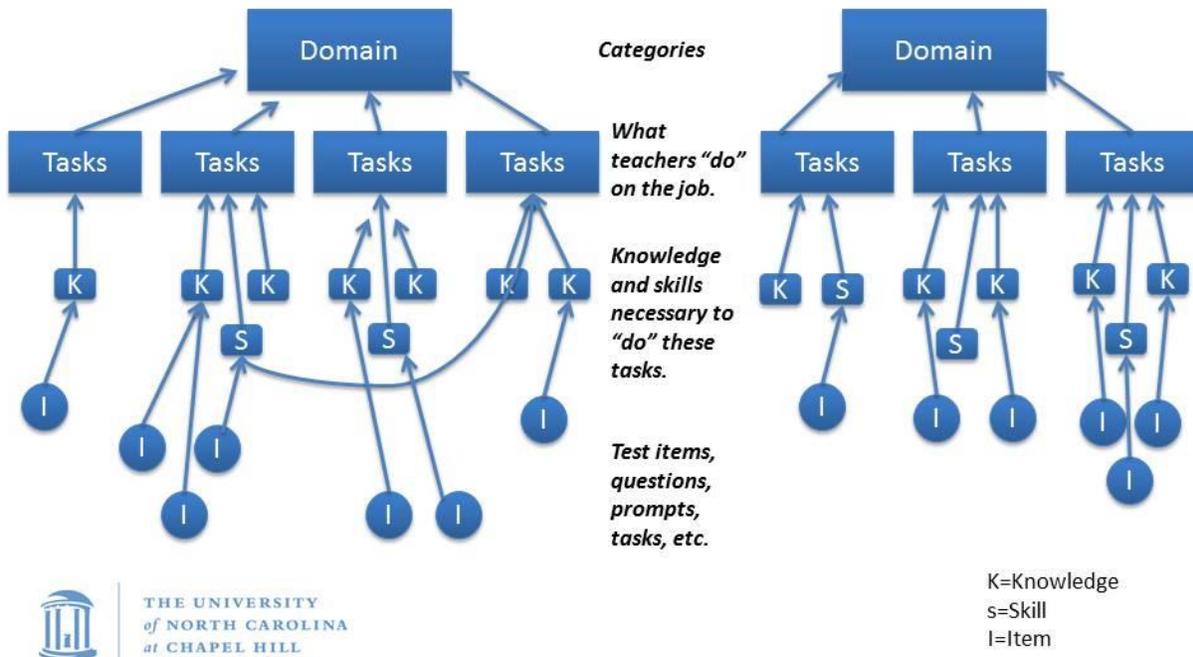
Math:

Number rods/Red and blue rods	Sandpaper numerals	Spindle Boxes
Cards and counters board	Golden Bead materials	Stamp Game/100
Bead Frames	Ten Board	Teen Board
Hundred Board	Addition Strip Board	Division Board
Subtraction Strip Board	Multiplication Board	

Cultural:

Globes	Land Forms	Maps
Cultural Geography	History	Plant puzzles
Botany 3 part cards	Animal puzzles	Zoology 3 part cards
Art materials	Library area	Music area

Domains, Tasks, Knowledge, Skills, and Items



Example of Domain, Task, KSA hierarchy

Domain	<u>The prepared environment:</u> A classroom environment that is developmentally appropriate for the physical, cognitive, and social-emotional growth and development of children in the mixed age group.
Task	Organize the Montessori environment for the mixed age group with developmentally appropriate Montessori materials in order to support the physical, cognitive, socio-emotional development of the students.
*Knowledge of	Each curriculum area (practical life, sensorial, math, language, cultural, science, physical and cultural geography)

	Sequence and order of lessons in each curriculum area
*Skill in	Designing an early childhood Montessori environment that orders the lessons and activities in each curriculum area

*There would typically be many of these statements for each task. This is just an example since we will not be drafting these statements in this meeting.

ECMT Certification Domains

These domains were initially developed by the advisory committee to provide the practice analysis panel with a starting point for categorizing essential tasks. We will review these briefly at the beginning of the process and then revise as we go through the process of developing and categorizing tasks.

- I. Individualized Instruction:** Providing instruction that is developmentally appropriate for each student in each curriculum area based on assessment of individual learning needs.

- II. Prepared environment:** The classroom environment that is developmentally appropriate for the physical, cognitive, and social-emotional development of the multi-aged group of students.

- III. Classroom/Behavior Management:** The use of management techniques, the Montessori Peace curriculum, lessons of Grace and Courtesy, and the preparation

of the environment to provide guidance and support for the social-emotional development of the mixed age group resulting in a respectful classroom community with positive social interactions.

IV. Assessment: Evaluating students using formative (i.e., observation, check lists) and summative assessments in order to teach to individual learning needs.

V. Communications/Interactions: Engagement in interactions with children, parents, co-workers, and professionals, using a variety of means and methods, in order to communicate, educate and collaborate in a positive manner that supports the developmental needs of each student in the multi-age group.

Identifying tasks

Close your eyes and imagine you are a new teacher working independently in a classroom for the first time. Some people do better thinking of an entry level teacher that they know. What are the critical things you must be able to do in order to be a *just qualified* teacher?

What is critical?

- Tasks that must be done competently in order to avoid harm to the public (students, parents, school community).

What is done most frequently?

- A task may be really important, but if it is only done occasionally it may not be essential to test.

What are the tasks that are specific to an early childhood Montessori teacher?

- Certification tests avoid general knowledge (i.e., intelligence, communication skills, general content knowledge).

What are the tasks that are measurable on a standardized test?

- Not measurable: personality/psychological traits, ethical behavior
- Measurable: knowledge and skills necessary to competently do critical tasks

❖ Small Group Activity

1. Identify a critical task.
2. Decide on the category or *Domain* with which it aligns.
3. Write the task on a post it of the color that aligns with the *Domain* and place it in the designated area. There is a place for tasks that do not align with any of the existing categories.

Domains

- I. Individualized Instruction - Yellow**
- II. Prepared environment - Magenta**
- III. Classroom/Behavior Management - Red**
- IV. Assessment – Dark Green**
- V. Communications/Interactions – Blue**
- VI. Other – Light Green**

Anatomy of a task statement

Once the critical tasks are identified and organized, we will revise these into formal task statements similar to the example on page 8.

Example: Organize the Montessori environment for the mixed age group with developmentally appropriate Montessori materials in order to support the physical, cognitive, socio-emotional development of the students.

- Verb – Organize
- Direct object – Organize what? the Montessori environment
- Using what tool(s) – developmentally appropriate materials
- For what purpose – to provide?

❖ Small group activity

Each group will be assigned one domain. In your groups, revise tasks in your assigned domain as formal task statements.

APPENDIX H. INITIAL TASKS DEVELOPED IN SMALL GROUPS

Individual Instruction

Revised domain name to Instruction

1. Observe the child
 2. Planning for individual instructions
 3. Understanding coordination, independents, concentration, order
 4. Planning for individual instruction
 5. Keep records
 6. Interpret record keeping
 7. Recognizing indications of sensitive periods of development
 8. Labeling the lesson
 9. Providing language to understand a concept
 10. Present individual and small group lessons
 11. Invite and entice child to lesson
 12. Present lessons
 13. Adapt lesson presentation according to individual needs
-

Classroom/Behavior Management

Revised domain name to Classroom Management

1. Observe the child
2. Observe/monitor classroom
3. Guide/redirect children
4. Redirecting behavior
5. Assess/determine when it is time to work with a child/children or give a lesson
6. Determine when it is appropriate to intervene
7. Guide/assess children in peaceful conflict resolution
8. Encourage independence
9. Greeting children at beginning of the day
10. Daily routine/schedule uninterrupted work cycles
11. Collaborate with assistant
12. Establish ground rules and limits
13. Manage transition times
14. Lead large group activities
15. Redirecting children as situation warrants
16. Modeling grace and courtesy x2
17. Use of grace and courtesy between teacher and student
18. Ground rules established grace/courtesy lesson
19. Create a procedure for conflict resolution
20. Awareness of states/behavior norms (false fatigue, developmental)
21. Visible signs of expected procedures (ground rules, bells, hand signals)
22. Identified structures of a peace curriculum

23. Understanding of establishing order (techniques to)

Prepared Environment

1. Sequencing lessons on shelf
 2. Create new lessons
 3. Organize materials
 4. Arrange materials in proper sequence
 5. Available ample individualized work spaces
 6. Repair materials
 7. Set up individual areas
 8. Arrange room to include all areas
 9. Sequencing materials
 10. Accessible materials
 11. Ensure that materials are sequentially organized on shelves
 12. Prepare a daily schedule
 13. Establish the flow of classroom
 14. Evaluate productivity
 15. Provide nature into the classroom with plants and animals
 16. Resetting the classroom
 17. Sequenced materials – not a task
 18. Establish defined areas
 19. Outdoor classroom – not a task
 20. Prepare and maintain individual works
-

Assessment

1. Observe the child
2. Organize a system for record keeping and planning
3. Organized record keeping system
4. Keep records
5. Interpret record keeping
6. Record observations and presentation/practice/mastery of work
7. Determine mastery
8. Plan next lesson for a given student

Communication/Interaction

Revised name of domain to Communication and Interaction

1. Conferencing with parents
2. Collaborate with assistant

3. Write a classroom newsletter
4. Coordinate tasks among adults in classroom
5. Direct/train assistants
6. Communicate/report to parents
7. Plan and implement parent education
8. Hold parent teacher conference
9. Seek opportunities for professional development
10. Effectively communicate with stakeholders
11. Collaborate with other teachers
12. Support intrinsic motivation
13. Discerns which intervention is appropriate
14. Works as a team member of other team members
15. Ability to tell class events
16. Use of technology
17. Understanding meanings of lessons

APPENDIX I. PANEL MEETING EVALUATION QUESTIONNAIRE

ECMT Practice Analysis Study Feedback

On a scale of 1 (*Disagree*) to 4 (*Strongly Agree*) please rate each of the following statements, and add any additional comments at the end.

	Disagree			Strongly Agree
The orientation provided me with a clear understanding of the purpose of the meeting.	1	2	3	4
The workshop leader clearly explained the tasks.	1	2	3	4
The training helped me to understand how to perform the task.	1	2	3	4
The large and small group discussions aided my understanding of the process.	1	2	3	4
There was equal opportunity for everyone to contribute ideas and opinions.	1	2	3	4
The preworkshop materials were helpful in understanding what my role and a practice analysis are.	1	2	3	4
I am confident about the defensibility and appropriateness of the Practice Analysis panel's work.	1	2	3	4

Please describe one or two *specific* changes that would help improve these workshops:

* Workshop feedback instrument developed by Dr. Stephen Johnson.

APPENDIX J. PILOT QUESTIONNAIRE

ECMT Certification Pilot Survey

Q1.1 Do you work in South Carolina?

- Yes (1)
- No (2)

Q1.2 Do you work in early childhood Montessori education (teacher, teacher preparation, administration)?

- Yes (1)
- No (2)

Q2.1 Thank you for your participation in this study! The following questionnaire will ask you to do five things. 1. Consider the definition of a "just qualified" early childhood Montessori teacher (ECMT) that was drafted for this study. "Just qualified" means that the teacher has the minimum knowledge and skills to competently perform the essential tasks necessary for teaching in an early childhood Montessori classroom. 2. Rate the presented tasks on two scales: criticality and frequency. These tasks were developed in a face-to-face meeting with a group of ECMTs, ECMT educators, and administrators representing the various practice settings (i.e., public and private Montessori classrooms) and four Montessori affiliates (AMI, AMS, MEPI, and Independent) in South Carolina. 3. Add any additional tasks you feel are not covered in each domain. When doing so, please keep in mind that a credentialing test cannot cover all tasks that an ECMT does. Our goal is to indicate the most critical and frequently performed tasks that are measurable in a standardized manner (precludes soft skills, personality traits, etc). 4. Answer a series of questions about your education, experience, and demographics intended to demonstrate that the

participants of this survey are representative of the ECMT profession. No individual identifications will be requested, nor any individual responses reported. School information is collected for the purpose of reporting response rates. 5. Since this is a pilot test of the questionnaire, it would be helpful to have your feedback on clarity and ease of use of questionnaire. Let's get started!

Q44 1. Definition of an Early Childhood Montessori Teacher (ECMT) A "just qualified ECMT" is defined for the purposes of this study as an entry-level teacher who has graduated from a MACTE accredited early childhood preparation program, which requires a minimum of 200 hours of academic work and 400 hours of practicum work. Teachers are prepared at a minimum in areas of philosophy, theory, child development, practical life, sensorial, mathematics, language, science, physical geography, cultural studies, physical education, music, and arts education. The ECMT has experience as a supervised teacher during the practicum portion of their education, but is not expected to have unsupervised experience in the ECM classroom. A just qualified ECMT will have knowledge of the basic materials in a Montessori classroom. In order to qualify to take a future ECMT test, teachers would need fulfill the following requirements: Ø Certification from a MACTE accredited teacher preparation program Ø Minimum 400 hours of supervised teaching Ø Bachelor's degree is recommended, but not required

Q2.2 Task ratings - Criticality Criticality for this questionnaire is the extent to which competent performance on a particular task is necessary for an ECMT. · If a task is not performed, or is performed incorrectly, would there be adverse consequences for a child, a group of children, the school and/or families? · Adverse consequences may be physical, mental,

emotional, or financial. For example, a child fails to learn, a child is physically, emotionally, mentally harmed, other members of the school community are impacted (i.e., another teacher has to compensate for lack of competency). For each task, rate the degree of risk on the following scale. If you choose 0, then you are indicating that there is no risk of adverse consequences if a teacher cannot competently perform that task. If you select 5, then you are indicating that there is a very high risk of adverse consequences if a teacher cannot perform that task competently.

No risk of adverse			Very high risk of		
consequences			adverse consequences		
0	1	2	3	4	5

Here's an example: Task: Provide a well sequenced art area with a wide variety of materials in order to promote independent learning. Ask yourself: How much risk of adverse consequences is there if an entry-level ECMT is not capable of completing this task? Rating: If you think the risk is high, then select a higher number. If you think there is little to no risk indicate a zero or low rating on the scale.

Q48 For each task, rate the degree of risk on the following scale. If you choose 0, then you are indicating that there is no risk of adverse consequences if a teacher cannot competently perform that task. If you select 5, then you are indicating that there is a very high risk of adverse consequences if a teacher cannot perform that task competently.

Q3.1 Domain I. Instruction Providing instruction that is developmentally appropriate for each student in each curriculum area based on assessment of individual learning needs.

	How critical is each task?						How frequently is each task performed?					
	0 No risk of adverse consequences (1)	1 (2)	2 (3)	3 (4)	4 (5)	5 Very high risk of adverse consequences (6)	0 Never performed (1)	1 (2)	2 (3)	3 (4)	4 (5)	5 Performed very often (6)
Present individual and small group lessons across all curriculum areas using Montessori materials and activities to promote acquisition of skills and/or concepts. (1)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Adapt lesson presentation and environment based on individual needs, including learning differences, as determined by observation in order to promote development. (2)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Q3.2 Are there other critical, frequently completed tasks related to instruction that you think should be included in this domain?

of the classroom environment using observation in order to meet the changing needs of the children. (8)													
---	--	--	--	--	--	--	--	--	--	--	--	--	--

<p>Redirect a child who is engaged in disruptive, destructive, or dangerous behavior in order to encourage engagement and normalization. (2)</p>	○))))	○	○))))	○
<p>Establish ground rules and limits in cooperation with the children, using lessons such as grace and courtesy and community building, in order to encourage engagement and normalization. (3)</p>	○))))	○	○))))	○
<p>Manage transition times using auditory and visual signals in order to encourage independence and normalization. (4)</p>	○))))	○	○))))	○
<p>Facilitate large group activities (i.e. games, storytelling, walking the line, silence game, music enrichment) in order to establish a classroom community. (5)</p>	○))))	○	○))))	○
<p>Create a classroom where</p>	○))))	○	○))))	○

Q6.1 Domain IV. Assessment Evaluating students using formative (i.e., observation, check lists) and summative assessments in order to teach to individual learning needs.

	How critical is each task?						How frequently is each task performed?					
	0 No risk of adverse consequences (1)	1 (2)	2 (3)	3 (4)	4 (5)	5 Very high risk of adverse consequences (6)	0 Never performed (1)	1 (2)	2 (3)	3 (4)	4 (5)	5 Performed very often (6)
Assess the child using observation (i.e., running records, anecdotal, time sampling, check lists) to evaluate interest, progress, and subsequent presentations. (1)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Maintain a system of record keeping and planning to document the progress of each child's presentation, practice, and mastery of skills and concepts. (2)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Q6.2 Are there other critical, frequently completed tasks related to assessment that you think should be included in this domain?

Q51 Criticality - Is there risk of adverse consequences (i.e., child does not learn, child is physically, emotionally, mentally harmed, harm to other members of the school community) if an entry-level ECMT does not perform the task competently?

No risk of adverse						Very high risk
of consequences						adverse consequences
0	1	2	3	4	5	

Frequency - How frequently does an entry-level ECMT perform the task?						Never
perform						Perform very often
0	1	2	3	4	5	

Q7.1 Domain V. Communication and Interaction Engagement in interactions with children, parents, co-workers, and professionals, using a variety of means and methods, in order to communicate, educate and collaborate in a positive manner that supports the developmental needs of each student in the multi-age group.

	How critical is each task?					How frequently is each task performed?						
	0 No risk of adverse consequences (1)	1 (2)	2 (3)	3 (4)	4 (5)	5 Very high risk of adverse consequences (6)	0 Never performed (1)	1 (2)	2 (3)	3 (4)	4 (5)	5 Performed very often (6)
Communicate with parents on a regular basis through a variety of means (i.e., parent/teacher conferences, parent education, newsletters, reports, email, phone) in order to encourage a seamless transition for the child between home and school. (1)	<input type="radio"/>					<input type="radio"/>	<input type="radio"/>					<input type="radio"/>
Establish a partnership with assistant(s) and/or co-teacher(s) using communication tools (i.e., team meetings, education, job descriptions, set expectations) in order to create a consistent and productive learning environment. (2)	<input type="radio"/>					<input type="radio"/>	<input type="radio"/>					<input type="radio"/>
Interact respectfully with each child	<input type="radio"/>					<input type="radio"/>	<input type="radio"/>					<input type="radio"/>

by creating and taking advantage of opportunities to listen, engage with, and/or communicate with in order to meet each child's individual needs. (3)															
---	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

Q7.2 Are there other critical, frequently completed tasks related to communication and interaction that you think should be included in this domain?

Q45 Demographic Questions

Q8.1 1. What is your sex?

Label: Sex

- Male (1)
- Female (2)

Q47 Do you consider yourself to be Hispanic/Latino?

Label: Ethnicity

- Yes (1)
- No (2)

Q8.3 Select one or more of the following racial categories to describe yourself?

Label: Race (V65)

- American Indian or Alaskan Native (1)
- Asian (2)
- Black or African American (3)
- Native Hawaiian or Pacific Islander (4)
- White (5)
- Other (7) _____

Q8.4 What is your age?

Label: Age

- Under 25 (1)
- 26-35 (2)
- 36-45 (3)
- 46-55 (4)
- Over 55 (5)

Q8.5 What is your highest level of education, not including your Montessori training?

Label: Education

- High School Diploma (1)
- Some college (2)
- Associate Degree (3)
- Bachelor's Degree (4)
- Master's Degree (5)
- Doctoral Degree (6)
- Other (7)

Q8.6 If education is other, please explain.

Q8.7 What was the affiliation of the Montessori teacher or administrator preparation program you attended? (If you attended more than one, list the affiliation of the program you graduate from or attended most recently.)

- Independent (IND) (7)
- American Montessori Association (AMS) (1)
- Pan American Montessori Society (PAMS) (6)
- Association Montessori International (AMI) (2)
- Montessori Education Programs International (MEPI) (3)
- International Association of Progressive Montessori (IAPM) (4)
- International Montessori Council (IMC)/Center for Guided Montessori Studies (CGMS) (5)
- I have not attended a Montessori preparation program. (8)

Q8.8 For what level(s) are you Montessori certified?

- Assistant (1)
- Primary/Early Childhood (2)
- Elementary I (3)
- Elementary II (4)
- Secondary I (5)
- Secondary II (6)
- Administration (8)
- Not certified (7)

Q8.9 Which best describes the area you work in?

- Urban (1)
- Suburban (2)
- Rural (3)

Q8.10 Which geographical region of South Carolina do you work in?

- Upstate (1)
- Midlands (2)
- Low Country (3)

Q8.11 Are you currently working as an early childhood Montessori teacher?

- Yes (1)
- No (2)

If Yes Is Selected, Then Skip To Which best describes the Montessori p...If No Is Selected, Then Skip To Are you currently working as an early...

Q8.12 Which best describes the Montessori practice setting in which you teach?

Label: Practice Setting

- Private Montessori school (1)
- Public Montessori magnet program (5)
- Public Montessori charter school (3)
- Public Montessori school (non-charter) (4)
- Public Montessori classroom within a traditional school (2)
- Other (9) _____

Q8.13 How many years experience do you have as an early childhood Montessori teacher?

Label: ECTeachExperience

- 0-3 years (1)
- 4-7 years (2)
- 8-11 years (3)
- 12-15 years (4)
- More than 15 years (5)

Q8.14 Are you a South Carolina state licensed teacher?

Label: Licensed

- Yes (1)
- No (2)

Q8.15 Are you currently working as an early childhood Montessori teacher educator?

Label: ECTeachEd

- Yes (1)
- No (2)

If Yes Is Selected, Then Skip To What was the affiliation of the Monte...If No Is Selected, Then Skip To Are you a Montessori school administr...

Q8.16 What is the affiliation of the Montessori teacher preparation program in which you teach?

Label: TeachEdAff

- Independent (IND) (7)
- American Montessori Association (AMS) (1)
- Pan American Montessori Society (PAMS) (6)
- Association Montessori International (AMI) (2)
- Montessori Education Programs International (MEPI) (3)
- International Association of Progressive Montessori (IAPM) (4)
- International Montessori Council (IMC)/Center for Guided Montessori Studies (CMGS) (5)

Q8.17 How many years of experience do you have as an early childhood Montessori teacher educator?

Label: TeachEdExp

- 0-3 years (1)
- 4-7 years (2)
- 8-11 years (3)
- 12-15 years (5)
- More than 15 years (6)

Q8.18 Are you an administrator (i.e., head of school, director, assistant director, principal, vice principal) of a Montessori school or program that has an early childhood classroom(s)?

Label: Administrator

- Yes (1)
- No (2)

If Yes Is Selected, Then Skip To How many years of experience do you h...If No Is Selected, Then Skip To In what school or training program do...

Q8.19 How many years of experience do you have in Montessori-related administration?

Label:AdminExp

- 0-3 years (1)
- 4-7 years (2)
- 8-11 years (3)
- 12-15 years (4)
- More than 15 years (5)

Q8.20 Which best describes the Montessori practice setting you work in?

Label:AdminPracticeSetting

- Private Montessori school (1)
- Public Montessori magnet program (5)
- Public Montessori charter school (3)
- Public Montessori school (non-charter) (4)
- Public Montessori classroom within a traditional school (2)
- Other (9) _____

Q8.21 In what school or training program do you currently work?

Q46 Pilot Questionnaire Evaluation

Q9.1 The purpose of this section is to assess the clarity and usability of the questionnaire. Please indicate your level of agreement with each statement.

	Strongly disagree 1 (8)	2 (2)	3 (3)	4 (4)	5 (5)	Strongly Agree 6 (12)
The purpose of the study was clearly stated. (1)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
The instructions for how to complete the questionnaire were clear. (4)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I understood how to rate each task using the criticality and frequency response formats. (2)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
The online questionnaire software (Qualtrics) was easy to use. (3)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Q9.2 Please describe any errors, typos, or issues that you noticed in the questionnaire. You may also provide general feedback related to the questionnaire.

APPENDIX K. EMAIL COMMUNICATIONS FOR SURVEY

Pilot Survey Invitation

Are you an early childhood Montessori teacher, teacher educator, or administrator? Be one of the first to share your expertise and participate in this pilot survey.

If you teach in an early childhood Montessori program, prepare early childhood Montessori teachers, or direct a Montessori program in South Carolina then I need your input on an important research study!

The purpose of this study is to take the first step in determining the content framework for a potential industry certification test for early childhood Montessori teachers (ECMTs). An industry certification is a credential that is independent of specific training programs or affiliates that aligns with the actual tasks that a teacher must perform competently. This first step would lead to identification of knowledge, skills, and abilities necessary to perform each task, followed by creation of test content for such a certification test.

This is an academic study being conducted for research purposes. However, the results of this study may be used at some point in the future to conduct a pilot program in South Carolina or to inform a national study.

Participants will be asked to respond to a brief (approximately 20 minute) questionnaire.

The questionnaire contains the essential tasks of an entry level ECMT, which you will be asked to rate based on their importance and performance frequency. These tasks were defined by a panel of 13 teachers, teacher educators, and administrators from across South Carolina, representing public and private Montessori programs. The group also represented graduates of and educators in AMI, AMS, MEPI, and Independent affiliated training programs.

Benefits of Participation are demonstration of leadership in your profession by contributing your expertise to the development of an industry certification and involvement in expanding the body of research related to Montessori education.

What is a pilot survey? In research, we typically administer a survey instrument, the questionnaire, to a small, representative sample in order to identify any issues prior to sending it out to the broader population. The only difference is that you will be asked to respond to 5 questions at the end regarding clarity and ease of use. Your responses are collected and used as part of the final study, so please answer all questions as accurately as possible.

To participate, simply click on the following link and begin the questionnaire!

https://unc.az1.qualtrics.com/SE/?SID=SV_eW2ZjZbkCPmopWR

The pilot phase of this study will end at midnight, Friday, January 22.

If you have any questions or trouble accessing the questionnaire please contact me at dtalley@live.unc.edu or by phone at 919-410-7792.

Thank you for your sharing your expertise for the benefit of research and the Montessori community!

Announcement for final survey

«GreetingLine»

Montessori-related research survey begins Monday, February 1!

School administrators, please share this announcement with your early childhood Montessori teachers!*

South Carolina Early childhood Montessori teachers, school administrators, and teacher educators, I need your expertise and experience! Please share your valuable insight on this important research study by completing a brief survey that will be launched February 1.

What: A research study to establish the content framework for an industry certification exam specific to early childhood Montessori teachers.

Why: A Certification is one indicator of professional competency. Most professions have such a certification; examples include traditional teachers, occupational therapists, speech therapists, psychologists, and nurses. *Currently the exams used to credential teachers are not Montessori-specific and, therefore, do not indicate the competency of a Montessori teacher.*

Benefits of an affiliation-independent certification:

- **For Montessori teachers-** New and experienced teachers in private and public schools would have a *portable measure of competency* to share with employers, as well as *increased professionalism*. Such a certification *increases the awareness and perception* of professionalism in the field, which could impact Montessori education in a powerful way.
- **For Montessori schools-** Both private and public schools would have one measure of a teacher's ability, independent of training programs and/or affiliations. Schools could use credentialed and independently certified teachers as a marketing tool.
- **For Montessori teacher education programs-** Provides an indicator of program quality in addition to other evaluation methods.
- **For Consumers-** Families enrolling in private and public schools would have an independent indicator of teacher competency.
- **For State Departments of Education and Licensing -** Provides a measure of accountability for recognition of the Montessori credential.

Who: A Panel of public and private South Carolina early childhood Montessori teachers, administrators, and teacher trainers have worked over the past two months with a testing expert on the initial phases of a professional practice analysis, which is a standard method used in the credentialing industry to identify the critical tasks performed by a professional – an Early Childhood Montessori Teacher.

Diane Talley, senior researcher, doctoral candidate, psychometrician, and Montessori mom will use data from this project to complete dissertation research, providing the foundation for a possible Montessori-specific certification exam.

How: Please participate in this final phase of the research study by completing a 15-20 minute questionnaire that will be available February 1-8. Your responses will:

- Verify the importance and frequency of each task
- Add your voice to the competencies needed by a beginning early childhood Montessori teacher
- Contribute to the development of an exam that could be used in SC and nationally

An email will be sent on February 1 with a link to this online questionnaire.

Thank you and feel free to contact me with any questions!

*Note for public school administrators – District approval was obtained prior to sending out this request.

Diane M. Talley

Doctoral Candidate

University of North Carolina, Chapel Hill

dtalley@live.unc.edu

919-410-7792

<https://www.linkedin.com/in/dianetalley>

APPENDIX L. FINAL QUESTIONNAIRE

ECMT Certification Questionnaire

Q1.1 Do you work in South Carolina?

- Yes
- No

Q1.2 Do you work in early childhood Montessori education (teacher, teacher education, administration)?

- Yes
- No

Q2.1 Thank you for your participation in this research study that is intended to establish the content framework for an industry certification exam specific to early childhood Montessori teachers. Remember that your responses are strictly confidential. Please read the instructions carefully! The following questionnaire will ask you to do four things. 1. Consider the definition of a "just qualified" early childhood Montessori teacher (ECMT) that was drafted for this study. "Just qualified" means that the teacher has the minimum knowledge and skills to competently perform the essential tasks necessary for teaching in an early childhood Montessori classroom. 2. Rate the presented tasks on two scales: criticality and frequency. These tasks were developed in a face-to-face meeting with a group of ECMTs, ECMT educators, and administrators representing the various practice settings (i.e., public and private Montessori classrooms) and four Montessori affiliates (AMI, AMS, MEPI, and Independent) in South Carolina. 3. Add any additional tasks you feel are not covered in each domain. When doing so, please keep in mind that a credentialing test cannot cover all tasks that an ECMT does. Our goal

is to indicate the most critical and frequently performed tasks that are measurable in a standardized manner (precludes soft skills, personality traits, etc). 4. Answer a series of questions about your education, experience, and demographics intended to demonstrate that the participants of this survey are representative of the ECMT profession in South Carolina. No individual identifications will be requested. School information is collected for the purpose of reporting response rates. Let's get started!

Q3.1 The Early Childhood Montessori Teacher (ECMT): An entry-level, or first year, Montessori teacher Completion of a training program accredited by the Montessori Accreditation Council for Teacher Education (MACTE) requiring a minimum of 200 academic hours (philosophy, child development, all curriculum areas, and classroom leadership) and a minimum of 400 hours mentored internship. May have a bachelor's degree, but not required The ECMT has experience as a supervised teacher during the practicum portion of their education, but is not expected to have unsupervised experience in the ECM classroom. A just qualified ECMT will have knowledge of the basic materials in a Montessori classroom.

Q4.1 Task Ratings Criticality for this questionnaire is the extent to which competent performance on a particular task is necessary for an ECMT in order to avoid adverse consequences. If a task is not performed, or is performed incorrectly, would there be adverse consequences for a child, a group of children, the school and/or families? Adverse consequences may be physical, mental, emotional, or financial. For example, a child fails to learn, a child is physically, emotionally, mentally harmed, other members of the school community are impacted (i.e., another teacher has to compensate for lack

of competency). For each task, rate the degree of risk on the following scale. If you choose 0, then you are indicating that there is no risk of adverse consequences if a teacher cannot competently perform that task. If you select 5, then you are indicating that there is a very high risk of adverse consequences if a teacher cannot perform that task competently.

No risk of adverse of consequences	Very high risk adverse consequences
0 1 2 3 4	5

Here's an example: Task: Provide a well sequenced art area with a wide variety of materials in order to promote independent learning. Ask yourself: How much risk of adverse consequences is there if an entry-level ECMT is not capable of completing this task? Rating: If you think the risk is high, then select a higher number. If you think there is little to no risk indicate a zero or low rating on the scale.

Q5.1 For each task, rate the degree of risk on the following scale. If you choose 0, then you are indicating that there is no risk of adverse consequences if a teacher cannot competently perform that task. If you select 5, then you are indicating that there is a very high risk of adverse consequences if a teacher cannot perform that task competently.

Q5.2 Domain I. Instruction Providing instruction that is developmentally appropriate for each student in each curriculum area based on assessment of individual learning needs.

	0	1	2	3	4	5
Present individual and small group lessons across all curriculum areas using Montessori materials and activities to promote acquisition of skills and/or	<input type="radio"/>					

concepts. Adapt lesson presentation and environment based on individual needs, including learning differences, as determined by observation in order to promote development.	<input type="radio"/>					
---	-----------------------	-----------------------	-----------------------	-----------------------	-----------------------	-----------------------

Q5.3 Are there other critical, frequently completed tasks related to instruction that you think should be included in this domain?

Q6.1 For each task, rate the degree of risk on the following scale. If you choose 0, then you are indicating that there is no risk of adverse consequences if a teacher cannot competently perform that task. If you select 5, then you are indicating that there is a very high risk of adverse consequences if a teacher cannot perform that task competently.

Q6.2 Domain II. Prepared Environment The classroom environment that is developmentally appropriate for the physical, cognitive, and social-emotional development of the multi-aged group of students.

	0	1	2	3	4	5
Establish procedures using tools such as a daily schedule, grace and courtesy lessons, expected procedures, ground rules, and demonstration of lesson to support independence in the child.	<input type="radio"/>					

<p>Sequence materials in each curriculum area using established Montessori guidelines to facilitate methodical acquisition of concepts and skills.</p>	○	○	○	○	○	○
<p>Create original activities using Montessori guidelines in order to meet the needs of children with specific learning styles and/or interests.</p>	○	○	○	○	○	○
<p>Maintain materials and the environment by repairing and restoring to ensure that all works are complete and ready for use by the children.</p>	○	○	○	○	○	○
<p>Prepare the environment using Montessori guidelines to create appropriate curriculum areas, work spaces and activities in order to foster movement, order and independence.</p>	○	○	○	○	○	○
<p>Assess the functionality of the classroom environment using observation in order to meet the changing needs of the children.</p>	○	○	○	○	○	○

Q6.3 Are there other critical, frequently completed tasks related to the prepared environment that you think should be included in this domain?

Q7.1 For each task, rate the degree of risk on the following scale. If you choose 0, then you are indicating that there is no risk of adverse consequences if a teacher cannot competently perform that task. If you select 5, then you are indicating that there is a very high risk of adverse consequences if a teacher cannot perform that task competently.

Q7.2 Domain III. Classroom Management The use of management techniques, the Montessori Peace curriculum, lessons of Grace and Courtesy, and the preparation of the environment to provide guidance and support for the social-emotional development of the mixed age group resulting in a respectful classroom community with positive social interactions.

	0	1	2	3	4	5
Guide a child to purposeful work through invitation to a lesson or activity in order to encourage engagement and normalization.	<input type="radio"/>					
Redirect a child who is engaged in disruptive, destructive, or dangerous behavior in order to encourage engagement and normalization.	<input type="radio"/>					
Establish ground rules and limits in cooperation with the children, using lessons such as grace and courtesy and community	<input type="radio"/>					

<p>building, in order to encourage engagement and normalization.</p>						
<p>Manage transition times using auditory and visual signals in order to encourage independence and normalization.</p>	○	○	○	○	○	○
<p>Facilitate large group activities (i.e. games, storytelling, walking the line, silence game, music enrichment) in order to establish a classroom community.</p>	○	○	○	○	○	○
<p>Create a classroom where children can resolve conflicts, using methods such as a peace curriculum, for the purpose of developing classroom community.</p>	○	○	○	○	○	○
<p>Assess the children's engagement level in the environment using observation in order to determine whether to begin interaction and/or lesson demonstration.</p>	○	○	○	○	○	○

Q7.3 Are there other critical, frequently completed tasks related to the classroom management that you think should be included in this domain?

Q8.1 For each task, rate the degree of risk on the following scale. If you choose 0, then you are indicating that there is no risk of adverse consequences if a teacher cannot competently perform that task. If you select 5, then you are indicating that there is a very high risk of adverse consequences if a teacher cannot perform that task competently.

Q8.2 Domain IV. Assessment Evaluating students using formative (i.e., observation, check lists) and summative assessments in order to teach to individual learning needs.

	0	1	2	3	4	5
Assess the child using observation (i.e., running records, anecdotal, time sampling, check lists) to evaluate interest, progress, and subsequent presentations.	<input type="radio"/>					
Maintain a system of record keeping and planning to document the progress of each child's presentation, practice, and mastery of skills and concepts.	<input type="radio"/>					

Q8.3 Are there other critical, frequently completed tasks related to the assessment that you think should be included in this domain?

Q9.1 For each task, rate the degree of risk on the following scale. If you choose 0, then you are indicating that there is no risk of adverse consequences if a teacher cannot competently perform that task. If you select 5, then you are indicating that there is a very high risk of adverse consequences if a teacher cannot perform that task competently.

Q9.2 Domain V. Communication and Interaction Engagement in interactions with children, parents, co-workers, and professionals, using a variety of means and methods, in order to communicate, educate and collaborate in a positive manner that supports the developmental needs of each student in the multi-age group.

	0	1	2	3	4	5
<p>Communicate with parents on a regular basis through a variety of means (i.e., parent/teacher conferences, parent education, newsletters, reports, email, phone) in order to encourage a seamless transition for the child between home and school.</p>	<input type="radio"/>					
<p>Interact respectfully with each child by creating and taking advantage of opportunities to listen, engage with, and/or communicate with in order to meet each child's individual needs.</p>	<input type="radio"/>					

Q9.3 Are there other critical, frequently completed tasks related to the communication and interaction that you think should be included in this domain?

Q10.1 You have finished rating all of the tasks for criticality. Next, please rate each task based on the frequency with which an entry-level ECMT would perform each task. Frequency - How frequently does an entry-level ECMT perform the task?

Never perform

Perform very

often 0 1 2 3 4 5

Q11.1 How frequently does an entry-level ECMT perform the task?

Q11.2 Domain I. Instruction Providing instruction that is developmentally appropriate for each student in each curriculum area based on assessment of individual learning needs.

	0	1	2	3	4	5
Present individual and small group lessons across all curriculum areas using Montessori materials and activities to promote acquisition of skills and/or concepts.	<input type="radio"/>					
Adapt lesson presentation and environment based on individual needs, including learning differences, as determined by observation in order to promote development.	<input type="radio"/>					

Q12.1 How frequently does an entry-level ECMT perform the task?

Q12.2 Domain II. Prepared Environment The classroom environment that is developmentally appropriate for the physical, cognitive, and social-emotional development of the multi-aged group of students.

	0	1	2	3	4	5
Establish procedures using tools such as a daily schedule, grace and courtesy lessons, expected procedures, ground rules, and	<input type="radio"/>					

<p>demonstration of lesson to support independence in the child.</p> <p>Sequence materials in each curriculum area using established Montessori guidelines to facilitate methodical acquisition of concepts and skills.</p> <p>Create original activities using Montessori guidelines in order to meet the needs of children with specific learning styles and/or interests.</p> <p>Maintain materials and the environment by repairing and restoring to ensure that all works are complete and ready for use by the children.</p> <p>Prepare the environment using Montessori guidelines to create appropriate curriculum areas, work spaces and activities in order to foster movement, order and independence.</p> <p>Assess the functionality of the classroom environment using</p>	○	○	○	○	○	○
--	---	---	---	---	---	---

observation in order to meet the changing needs of the children.						
--	--	--	--	--	--	--

Q13.1 How frequently does an entry-level ECMT perform the task?

Q13.2 Domain III. Classroom Management The use of management techniques, the Montessori Peace curriculum, lessons of Grace and Courtesy, and the preparation of the environment to provide guidance and support for the social-emotional development of the mixed age group resulting in a respectful classroom community with positive social interactions.

	0	1	2	3	4	5
Guide a child to purposeful work through invitation to a lesson or activity in order to encourage engagement and normalization.	<input type="radio"/>					
Redirect a child who is engaged in disruptive, destructive, or dangerous behavior in order to encourage engagement and normalization.	<input type="radio"/>					
Establish ground rules and limits in cooperation with the children, using lessons such as grace and courtesy and community building, in order to encourage engagement and normalization.	<input type="radio"/>					
Manage transition times using auditory and visual signals in order to encourage independence	<input type="radio"/>					

and normalization.						
Facilitate large group activities (i.e. games, storytelling, walking the line, silence game, music enrichment) in order to establish a classroom community.	<input type="radio"/>					
Create a classroom where children can resolve conflicts, using methods such as a peace curriculum, for the purpose of developing classroom community.	<input type="radio"/>					
Assess the children's engagement level in the environment using observation in order to determine whether to begin interaction and/or lesson demonstration.	<input type="radio"/>					

Q14.1 How frequently does an entry-level ECMT perform the task?

Q14.2 Domain IV. Assessment Evaluating students using formative (i.e., observation, check lists) and summative assessments in order to teach to individual learning needs.

	0	1	2	3	4	5
Assess the child using observation (i.e., running records, anecdotal, time sampling, check lists) to	<input type="radio"/>					

<p>evaluate interest, progress, and subsequent presentations.</p> <p>Maintain a system of record keeping and planning to document the progress of each child's presentation, practice, and mastery of skills and concepts.</p>	<input type="radio"/>					
--	-----------------------	-----------------------	-----------------------	-----------------------	-----------------------	-----------------------

Q15.1 How frequently does an entry-level ECMT perform the task?

Q15.2 Domain V. Communication and Interaction Engagement in interactions with children, parents, co-workers, and professionals, using a variety of means and methods, in order to communicate, educate and collaborate in a positive manner that supports the developmental needs of each student in the multi-age group.

	0	1	2	3	4	5
<p>Communicate with parents on a regular basis through a variety of means (i.e., parent/teacher conferences, parent education, newsletters, reports, email, phone) in order to encourage a seamless transition for the child between home and school.</p>	<input type="radio"/>					
<p>Interact respectfully with each child by creating and taking advantage of opportunities to listen, engage with, and/or communicate with in order to meet each child's individual needs.</p>	<input type="radio"/>					

Q16.1 Demographic Questions

Q16.2 1. What is your sex?

- Male
- Female

Q16.3 Do you consider yourself to be Hispanic/Latino?

- Yes
- No

Q16.4 Select one or more of the following racial categories to describe yourself?

- American Indian or Alaskan Native
- Asian
- Black or African American
- Native Hawaiian or Pacific Islander
- White
- Other _____

Q16.5 What is your age?

- Under 25
- 26-35
- 36-45
- 46-55
- Over 55

Q16.6 What is your highest level of education, not including your Montessori training?

- High School Diploma
- Some college
- Associate Degree
- Bachelor's Degree
- Master's Degree
- Doctoral Degree
- Other

Q16.7 If education is other, please explain.

Q16.8 What was the affiliation of the Montessori teacher or administrator preparation program you attended? (If you attended more than one, list the affiliation of the program you graduate from or attended most recently.)

- Independent (IND)
- American Montessori Association (AMS)
- Pan American Montessori Society (PAMS)
- Association Montessori International (AMI)
- Montessori Education Programs International (MEPI)
- International Association of Progressive Montessori (IAPM)
- International Montessori Council (IMC)/Center for Guided Montessori Studies (CGMS)
- I have not attended a Montessori preparation program.

Q16.9 For what level(s) are you Montessori certified?

- Assistant
- Primary/Early Childhood
- Elementary I
- Elementary II
- Secondary I
- Secondary II
- Administration
- Not certified

Q16.10 Which best describes the area you work in?

- Urban
- Suburban
- Rural

Q16.11 Which geographical region of South Carolina do you work in?

- Upstate
- Midlands
- Low Country

Q16.12 Are you currently working as an early childhood Montessori teacher?

- Yes
- No

If Yes Is Selected, Then Skip To Which best describes the Montessori p...If No Is Selected, Then Skip To Are you currently working as an early...

Q16.13 Which best describes the Montessori practice setting in which you teach?

- Private Montessori school
- Public Montessori magnet program
- Public Montessori charter school
- Public Montessori school (non-charter)
- Public Montessori classroom within a traditional school
- Other _____

Q16.14 How many years experience do you have as an early childhood Montessori teacher?

- 0-3 years
- 4-7 years
- 8-11 years
- 12-15 years
- More than 15 years

Q16.15 Are you a South Carolina state licensed teacher?

- Yes
- No

Q16.16 Are you currently working as an early childhood Montessori teacher educator?

- Yes
- No

If Yes Is Selected, Then Skip To What was the affiliation of the Monte...If No Is Selected, Then Skip To Are you a Montessori school administr...

Q16.17 What is the affiliation of the Montessori teacher preparation program in which you teach?

- Independent (IND)
- American Montessori Association (AMS)
- Pan American Montessori Society (PAMS)
- Association Montessori International (AMI)
- Montessori Education Programs International (MEPI)
- International Association of Progressive Montessori (IAPM)
- International Montessori Council (IMC)/Center for Guided Montessori Studies (CMGS)

Q16.18 How many years of experience do you have as an early childhood Montessori teacher educator?

- 0-3 years
- 4-7 years
- 8-11 years
- 12-15 years
- More than 15 years

Q16.19 Are you an administrator (i.e., head of school, director, assistant director, principal, vice principal) of a Montessori school or program that has an early childhood classroom(s)?

- Yes
- No

If Yes Is Selected, Then Skip To How many years of experience do you h...If No Is Selected, Then Skip To In what school or training program do...

Q16.20 How many years of experience do you have in Montessori-related administration?

- 0-3 years
- 4-7 years
- 8-11 years
- 12-15 years
- More than 15 years

Q16.21 Which best describes the Montessori practice setting you work in?

- Private Montessori school
- Public Montessori magnet program
- Public Montessori charter school
- Public Montessori school (non-charter)
- Public Montessori classroom within a traditional school
- Other _____

Q16.22 In what school or training program do you currently work?

APPENDIX M. EMAIL INVITATION FOR FINAL SURVEY

«GreetingLine»

South Carolina Early childhood Montessori teachers, school administrators, and teacher educators, I need your expertise and experience today! It's very easy. All you have to do to begin is click this link and **please read the instructions carefully!**

https://unc.az1.qualtrics.com/SE/?SID=SV_8J2bm6ekYjgRFfD

The questionnaire will be open through Monday, February 8. Please direct any questions or technical problems to dtalley@live.unc.edu.

Please share this link with any early childhood Montessori teachers, teacher educators, or administrators of early childhood Montessori programs in South Carolina!

Summary of the study:

What: A research study to establish the content framework for an industry certification exam specific to early childhood Montessori teachers.

Why: A Certification is one indicator of professional competency. Most professions have such a certification; examples include traditional teachers, occupational therapists, speech therapists, psychologists, and nurses. *Currently the exams used to credential teachers are not Montessori-specific and, therefore, do not indicate the competency of a Montessori teacher.*

Who: A Panel of public and private South Carolina early childhood Montessori teachers, administrators, and teacher trainers have worked over the past two months with a testing expert on the initial phases of a professional practice analysis, which is a standard method used in the credentialing industry to identify the critical tasks performed by a professional – an Early Childhood Montessori Teacher.

Diane Talley, senior researcher, doctoral candidate, psychometrician, and Montessori mom will use data from this project to complete dissertation research, providing the foundation for a possible Montessori-specific certification exam.

How: Please participate in this final phase of the research study by completing a 15-20 minute questionnaire. Your responses will:

- Verify the importance and frequency of each task
- Add your voice to the competencies needed by a beginning early childhood Montessori teacher
- Contribute to the development of an exam that could be used in SC and nationally

Thank you for sharing your valuable expertise!

Survey reminder

«GreetingLine»

A friendly reminder: You have until February 8th at 10:00pm to contribute your expertise!

Just click here: https://unc.az1.qualtrics.com/SE/?SID=SV_8J2bm6ekYjgRFfD

If you don't have access to a personal computer, the survey runs on tablets and smart phones too!

Please consider taking 10 minutes to support a research study that can benefit Montessori teachers in your state.

WHY?

- Currently there is not Montessori-specific credentialing test. This study provides the foundation for developing such a test for early child Montessori teachers in SC or the basis for a national study.
- A Montessori-specific certification test supports:
 - The need for competent teachers in every Montessori classroom
 - Professionalism of Montessori educators
 - Could lead to a path for recognition in every state

HOW?

- **Complete the questionnaire** to give your voice in defining the essential tasks of the early childhood Montessori teacher.
- The more responses, the more powerful our voice as a Montessori community

Join me in this valuable community effort.

Click here to begin: https://unc.az1.qualtrics.com/SE/?SID=SV_8J2bm6ekYjgRFfD

And thank you for your contribution to this research study and Montessori education!

APPENDIX N. METHODS COMPARISON FOR COMPUTING TASK WEIGHTS USING SURVEY RESPONSES

Task	[[C+1]*F] ^{1/2}			[(C+1) ² *F] ^{1/3}			C+F			5*(C-1)+F		
	<i>b</i>	<i>se</i>	<i>pwt</i>	<i>b</i>	<i>se</i>	<i>pwt</i>	<i>b</i>	<i>se</i>	<i>pwt</i>	<i>b</i>	<i>se</i>	<i>pwt</i>
D1T1	-1.73	0.5	8.14	-1.37	0.57	7.10	-1.41	0.4	8.20	0.16	0.07	4.64
D1T2	-0.38	0.4	5.90	-0.1	0.46	5.40	-0.22	0.24	5.72	-0.18	0.09	5.97
D2T1	-0.08	0.39	5.40	-0.1	0.46	5.40	-0.01	0.23	5.28	-0.25	0.1	6.25
D2T2	0.62	0.37	4.23	0.84	0.41	4.14	0.38	0.21	4.47	-0.03	0.08	5.38
D2T3	2.16	0.36	1.67	2.94	0.37	1.34	1.53	0.21	2.08	0.34	0.06	3.93
D2T4	-0.71	0.41	6.44	-0.31	0.47	5.68	-0.39	0.25	6.07	-0.02	0.07	5.34
D2T5	1.4	0.36	2.93	1.34	0.4	3.47	0.69	0.21	3.83	0.07	0.07	4.99
D2T6	0.76	0.36	4.00	1.01	0.41	3.92	0.38	0.21	4.47	0.1	0.07	4.87
D3T1	-0.89	0.43	6.74	-0.54	0.49	5.99	-0.59	0.27	6.49	-0.11	0.08	5.70
D3T2	-0.89	0.43	6.74	-1.37	0.57	7.10	-0.59	0.27	6.49	-0.31	0.1	6.48
D3T3	0.21	0.38	4.91	-0.54	0.49	5.99	0.04	0.22	5.18	-0.24	0.09	6.21
D3T4	0.07	0.38	5.15	0.84	0.41	4.14	0.04	0.22	5.18	0.14	0.07	4.72
D3T5	0.07	0.38	5.15	1.01	0.41	3.92	0.19	0.22	4.87	0.15	0.07	4.68
D3T6	0.76	0.36	4.00	1.18	0.4	3.69	0.56	0.21	4.10	0.1	0.07	4.87
D3T7	0.21	0.38	4.91	0.67	0.42	4.37	0.14	0.22	4.97	0.07	0.07	4.99
D4T1	-0.71	0.41	6.44	-1.37	0.57	7.10	-0.46	0.25	6.22	-0.03	0.08	5.38
D4T2	-0.55	0.41	6.18	-1.37	0.57	7.10	-0.22	0.24	5.72	-0.03	0.08	5.38
D5T1	1.4	0.36	2.93	-1.37	0.57	7.10	0.86	0.21	3.47	0.25	0.06	4.28
D5T2	-1.73	0.5	8.14	-1.37	0.57	7.10	-0.92	0.31	7.18	-0.17	0.09	5.93

**APPENDIX O. RELIABILITY ANALYSIS ACROSS METHODS FOR ESTIMATING
TASK WEIGHTS**

	$[(C+1)*F]^{1/2}$	$[(C+1)2*F]^{1/3}$	C+F	$5*(C-1)+F$
Panel				
<i>KR20</i>	0.9	0.91	0.91	0.93
Item Reliability	0.82	0.82	0.84	.77
Separation	2.15	2.31	2.48	2.01
Model <i>se</i>	0.4	0.47	0.24	0.47
Fit at mean	0.98/0.98	0.94/0.99	0.99/0.97	1.03/0.99
Survey				
<i>KR20</i>	0.96	0.96	0.97	0.97
Item Reliability	0.88	0.80	0.89	0.91
Separation	2.72	2.15	2.80	2.14
Model <i>se</i>	0.20	0.19	0.13	0.04
Fit at mean	1/0.97	1.04/0.98	0.99/0.98	1.05/0.97

APPENDIX P. BLUEPRINT WEIGHT COMPARISON

Task	Survey all			Survey - No extremes			Survey - No Administrators			Survey- No non-certified participants		
	<i>b</i> (<i>n</i> =84)	<i>se</i> (<i>n</i> =84)	% weights (<i>n</i> =84)	<i>b</i> (<i>n</i> =79)	<i>se</i> (<i>n</i> =79)	% weights (<i>n</i> =79)	<i>b</i> (<i>n</i> =72)	<i>se</i> (<i>n</i> =72)	% weights (<i>n</i> =72)	<i>b</i> (<i>n</i> =68)	<i>se</i> (<i>n</i> =69)	% weights (<i>n</i> =69)
D1T1	-0.14	0.19	6.10	-0.18	0.2	6.11	-0.17	0.21	6.19	-0.27	0.21	6.42
D1T2	-0.07	0.19	5.12	0.05	0.19	5.02	-0.12	0.21	5.34	-0.1	0.21	5.24
D2T1	-0.37	0.2	6.02	-0.42	0.2	6.04	-0.39	0.21	6.29	-0.5	0.22	6.42
D2T2	0.1	0.18	4.72	0.08	0.19	4.80	0.09	0.2	4.84	0.19	0.2	4.61
D2T3	1.23	0.16	1.60	1.25	0.16	1.85	1.2	0.17	1.80	1.34	0.17	1.48
D2T4	0.46	0.18	4.34	0.45	0.18	4.45	0.43	0.19	4.45	0.38	0.19	4.51
D2T5	0.07	0.19	4.88	0.05	0.19	4.95	0.21	0.2	4.55	0.15	0.2	4.61
D2T6	0.14	0.18	4.72	0.12	0.19	4.80	0.13	0.2	4.84	0.11	0.2	4.78
D3T1	0.14	0.18	5.28	0.12	0.19	5.27	0.21	0.2	5.24	0.15	0.2	5.24
D3T2	-0.94	0.21	7.54	-0.82	0.22	7.15	-0.86	0.22	7.21	-0.85	0.23	7.36
D3T3	-0.41	0.2	5.84	-0.46	0.21	5.87	-0.39	0.21	5.75	-0.55	0.22	6.02
D3T4	0.03	0.19	5.28	0.08	0.19	5.27	0.04	0.2	5.24	0.11	0.2	5.14
D3T5	0.14	0.18	5.12	0.12	0.19	5.19	0.04	0.2	5.44	-0.01	0.2	5.52
D3T6	-0.18	0.19	5.60	-0.14	0.2	5.56	-0.12	0.21	5.24	-0.1	0.21	5.42
D3T7	0.17	0.18	4.88	0.15	0.19	4.95	0.32	0.2	4.45	0.19	0.2	4.78
D4T1	-0.11	0.19	5.60	-0.14	0.2	5.56	-0.17	0.21	5.75	0.03	0.2	5.24
D4T2	-0.18	0.19	5.94	-0.14	0.2	5.80	-0.34	0.21	6.19	-0.06	0.2	5.82
D5T1	0.61	0.17	3.96	0.61	0.17	4.03	0.51	0.19	3.97	0.62	0.19	3.91
D5T2	-0.68	0.2	7.44	-0.77	0.22	7.32	-0.62	0.22	7.21	-0.8	0.23	7.48

Note. *Assuming a 100 item test and that number of test questions per task is equivalent to the rounded percentage.

**APPENDIX Q. FREQUENCY TABLES WITH PANEL AND SURVEY GROUP
RESPONSES TO CRITICALITY AND FREQUENCY SCALES FOR EACH TASK**

Criticality Scale

Panel D1T1CR		
	Frequency	Percent
4	5	31.3
5	11	68.8
Total	16	100.0

Survey D1T1CR		
	Frequency	Percent
0	1	1.3
2	6	7.6
3	6	7.6
4	25	31.6
5	41	51.9
Total	79	100

Panel D1T2CR		
	Frequency	Percent
3	1	6.3
4	4	25.0
5	11	68.8
Total	16	100.0

Survey D1T2CR		
	Frequency	Percent
0	1	1.3
1	1	1.3
2	4	5.1
3	9	11.4
4	24	30.4
5	40	50.6
Total	79	100

Panel D2T1CR

	Frequency	Percent
4	3	18.8
5	13	81.3
Total	16	100.0

Survey D2T1CR

	Frequency	Percent
0	1	1.3
2	6	7.6
3	5	6.3
4	25	31.6
5	42	53.2
Total	79	100

Panel D2T2CR

	Frequency	Percent
3	2	12.5
4	5	31.3
5	9	56.3
Total	16	100.0

Survey D2T2CR

	Frequency	Percent
0	1	1.3
1	2	2.5
2	5	6.3
3	10	12.7
4	29	36.7
5	32	40.5
Total	79	100

Panel D2T3CR		
	Frequency	Percent
1	1	6.3
2	1	6.3
3	4	25.0
4	8	50.0
5	2	12.5
Total	16	100.0

Survey D2T3CR		
	Frequency	Percent
0	1	1.3
1	3	3.8
2	9	11.4
3	20	25.3
4	26	32.9
5	20	25.3
Total	79	100

Panel D2T4CR		
	Frequency	Percent
3	4	25.0
4	5	31.3
5	7	43.8
Total	16	100.0

Survey D2T4CR		
	Frequency	Percent
0	1	1.3
1	2	2.5
2	5	6.3
3	10	12.7
4	38	48.1
5	23	29.1
Total	79	100

Panel D2T5CR		
	Frequency	Percent
3	2	12.5
4	8	50.0
5	6	37.5
Total	16	100.0

Survey D2T5CR		
	Frequency	Percent
0	1	1.3
1	1	1.3
2	7	8.9
3	4	5.1
4	33	41.8
5	33	41.8
Total	79	100

Panel D2T6CR		
	Frequency	Percent
1	1	6.3
3	3	18.8
4	5	31.3
5	7	43.8
Total	16	100.0

Survey D2T6CR		
	Frequency	Percent
0	1	1.3
1	1	1.3
2	6	7.6
3	9	11.4
4	29	36.7
5	33	41.8
Total	79	100

D3T1CR		
	Frequency	Percent
3	2	12.5
4	6	37.5
5	8	50.0
Total	16	100.0

Survey D3T1CR		
	Frequency	Percent
1	3	3.8
2	3	3.8
3	13	16.5
4	30	38
5	30	38
Total	79	100

D3T2CR		
	Frequency	Percent
4	4	25.0
5	12	75.0
Total	16	100.0

Survey D3T2CR		
	Frequency	Percent
1	2	2.5
2	3	3.8
3	8	10.1
4	18	22.8
5	48	60.8
Total	79	100

D3T3CR		
	Frequency	Percent
3	1	6.3
4	1	6.3
5	14	87.5
Total	16	100.0

Survey D3T3CR		
	Frequency	Percent
1	3	3.8
2	4	5.1
3	6	7.6
4	23	29.1
5	43	54.4
Total	79	100

Panel D3T4CR		
	Frequency	Percent
2	2	12.5
3	3	18.8
4	7	43.8
5	4	25
Total	16	100.0

Survey D3T4CR		
	Frequency	Percent
1	2	2.5
2	3	3.8
3	14	17.7
4	34	43
5	26	32.9
Total	79	100

Panel D3T5CR		
	Frequency	Percent
2	4	25
3	2	12.5
4	3	18.8
5	7	43.8
Total	16	100.0

Survey D3T5CR		
	Frequency	Percent
1	3	3.8
2	4	5.1
3	15	19
4	27	34.2
5	30	38
Total	79	100

Panel D3T6CR		
	Frequency	Percent
2	2	12.5
3	3	18.8
4	2	12.5
5	9	56.3
Total	16	100.0

Survey D3T6CR		
	Frequency	Percent
1	3	3.8
2	4	5.1
3	9	11.4
4	27	34.2
5	36	45.6
Total	79	100

Panel D3T7CR		
	Frequency	Percent
2	3	18.8
4	6	37.5
5	7	43.8
Total	16	100.0

Survey D3T7CR		
	Frequency	Percent
1	4	5.1
2	4	5.1
3	11	13.9
4	30	38
5	30	38
Total	79	100

Panel D4T1CR		
	Frequency	Percent
2	2	12.5
3	2	12.5
4	3	18.8
5	9	56.3
Total	16	100.0

Survey D4T1CR		
	Frequency	Percent
1	1	1.3
2	3	3.8
3	12	15.2
4	28	35.4
5	35	44.3
Total	79	100

Panel D4T2CR		
	Frequency	Percent
2	1	6.3
3	3	18.8
4	3	18.8
5	9	56.3
Total	16	100.0

Survey D4T2CR		
	Frequency	Percent
1	1	1.3
2	2	2.5
3	12	15.2
4	28	35.4
5	36	45.6
Total	79	100

Panel D5T1CR		
	Frequency	Percent
2	2	12.5
3	5	31.3
4	5	31.3
5	4	25
Total	16	100.0

Survey D5T1CR		
	Frequency	Percent
1	2	2.5
2	4	5.1
3	18	22.8
4	25	31.6
5	30	38
Total	79	100

D5T2CR

	Frequency	Percent
2	1	6.3
3	2	12.5
4	2	12.5
5	11	68.8
Total	16	100.0

Survey D5T2CR

	Frequency	Percent
1	1	1.3
2	1	1.3
3	8	10.1
4	26	32.9
5	43	54.4
Total	79	100

Frequency Scale

Panel D1T1FR

	Frequency	Percent
4	1	6.3
5	15	93.8
Total	16	100.0

Survey D1T1FR

	Frequency	Percent
1	1	1.3
2	2	2.5
3	8	10.1
4	12	15.2
5	56	70.9
Total	79	100.0

Panel D1T2FR

	Frequency	Percent
3	5	31.3
4	4	25.0
5	7	43.8
Total	16	100.0

Survey D1T2FR

	Frequency	Percent
1	1	1.3
2	6	7.6
3	12	15.2
4	20	25.3
5	40	50.6
Total	79	100.0

Panel D2T2FR

	Frequency	Percent
1	1	6.3
2	3	18.8
3	4	25.0
4	2	12.5
5	6	37.5
Total	16	100.0

Survey D2T2FR

	Frequency	Percent
1	1	1.3
2	2	2.5
3	5	6.3
4	23	29.1
5	48	60.8
Total	79	100.0

Panel D2T3FR

	Frequency	Percent
1	1	6.3
2	7	43.8
3	3	18.8
4	4	25.0
5	1	6.3
Total	16	100.0

Survey D2T3FR

	Frequency	Percent
1	2	2.5
2	9	11.4
3	20	25.3
4	24	30.4
5	24	30.4
Total	79	100.0

Panel D2T4FR

	Frequency	Percent
3	2	12.5
5	14	87.5
Total	16	100.0

Survey D2T4FR

	Frequency	Percent
1	1	1.3
2	4	5.1
3	8	10.1
4	18	22.8
5	48	60.8
Total	79	100.0

Panel D2T5FR

	Frequency	Percent
1	2	12.5
2	4	25.0
3	2	12.5
4	3	18.8
5	5	31.3
Total	16	100.0

Survey D2T5FR

	Frequency	Percent
1	1	1.3
2	4	5.1
3	5	6.3
4	19	24.1
5	50	63.3
Total	79	100.0

Panel D2T6FR

	Frequency	Percent
2	1	6.3
3	3	18.8
4	8	50.0
5	4	25.0
Total	16	100.0

Survey D2T6FR

	Frequency	Percent
1	1	1.3
2	3	3.8
3	10	12.7
4	20	25.3
5	45	57.0
Total	79	100.0

Panel D3T1FR

	Frequency	Percent
4	4	25.0
5	12	75.0
Total	16	100.0

Survey D3T1FR

	Frequency	Percent
1	1	1.3
2	2	2.5
3	4	5.1
4	18	22.8
5	54	68.4
Total	79	100.0

Panel D3T2FR

	Frequency	Percent
2	1	6.3
3	1	6.3
4	5	31.3
5	9	56.3
Total	16	100.0

Survey D3T2FR

	Frequency	Percent
2	4	5.1
3	4	5.1
4	16	20.3
5	55	69.6
Total	79	100.0

Panel D3T3FR

	Frequency	Percent
2	5	31.3
3	1	6.3
4	5	31.3
5	5	31.3
Total	16	100.0

Survey D3T3FR

	Frequency	Percent
1	1	1.3
2	4	5.1
3	6	7.6
4	16	20.3
5	52	65.8
Total	79	100.0

Panel D3T4FR

	Frequency	Percent
4	6	37.5
5	10	62.5
Total	16	100.0

Survey D3T4FR

	Frequency	Percent
1	1	1.3
2	3	3.8
3	6	7.6
4	16	20.3
5	53	67.1
Total	79	100.0

Panel D3T5FR

	Frequency	Percent
3	3	18.8
4	3	18.8
5	10	62.5
Total	16	100.0

Survey D3T5FR

	Frequency	Percent
1	1	1.3
2	4	5.1
3	7	8.9
4	16	20.3
5	51	64.6
Total	79	100.0

Panel D3T6FR

	Frequency	Percent
2	2	12.5
3	6	37.5
4	4	25.0
5	4	25.0
Total	16	100.0

Survey D3T6FR

	Frequency	Percent
2	5	6.3
3	6	7.6
4	20	25.3
5	48	60.8
Total	79	100.0

Panel D3T7FR

	Frequency	Percent
2	1	6.3
3	2	12.5
4	5	31.3
5	8	50.0
Total	16	100.0

Survey D3T7FR

	Frequency	Percent
1	2	2.5
2	2	2.5
3	8	10.1
4	14	17.7
5	53	67.1
Total	79	100.0

Panel D4T1FR

	Frequency	Percent
4	3	18.8
5	13	81.3
Total	16	100.0

Survey D4T1FR

	Frequency	Percent
0	1	1.3
1	1	1.3
2	1	1.3
3	9	11.4
4	22	27.8
5	45	57.0
Total	79	100.0

Panel D4T2FR

	Frequency	Percent
2	1	6.3
3	2	12.5
4	1	6.3
5	12	75.0
Total	16	100.0

Survey D4T2FR

	Frequency	Percent
0	1	1.3
1	1	1.3
2	1	1.3
3	8	10.1
4	19	24.1
5	49	62.0
Total	79	100.0

Panel D5T1FR

	Frequency	Percent
3	10	62.5
4	2	12.5
5	4	25.0
Total	16	100.0

Survey D5T1FR

	Frequency	Percent
0	2	2.5
2	5	6.3
3	13	16.5
4	22	27.8
5	37	46.8
Total	79	100.0

Panel D5T2FR

	Frequency	Percent
4	1	6.3
5	15	93.8
Total	16	100.0

Survey D5T2FR

	Frequency	Percent
1	1	1.3
2	3	3.8
3	3	3.8
4	15	19.0
5	57	72.2
Total	79	100.0

**APPENDIX R. FREQUENCY TABLES WITH PANEL AND SURVEY GROUP
RESPONSES TO COMBINED SCALE FOR EACH TASK**

Panel D1T1

	Frequency	Percent
5.48	6	37.5
6.00	10	62.5
Total	16	100.0

Survey D1T1

	Frequency	Percent
2.45	1	1.3
3.00	1	1.3
3.16	1	1.3
3.46	3	3.8
3.87	1	1.3
4.00	1	1.3
4.24	2	2.5
4.47	5	6.3
4.90	4	5.1
5.00	4	5.1
5.48	22	27.8
6.00	34	43.0
Total	79	100.0

Panel D1T2

	Frequency	Percent
4.47	2	12.5
4.90	4	25.0
5.00	1	6.3
5.48	4	25.0
6.00	5	31.3
Total	16	100.0

Survey D1T2

	Frequency	Percent
2.45	1	1.3
3.16	2	2.5
3.46	2	2.5
3.87	4	5.1
4.00	2	2.5
4.24	4	5.1
4.47	8	10.1
4.90	7	8.9
5.00	8	10.1
5.48	14	17.7
6.00	27	34.2
Total	79	100.0

Panel D2T1

	Frequency	Percent
3.46	1	6.3
3.87	1	6.3
4.24	2	12.5
4.90	2	12.5
5.00	2	12.5
5.48	2	12.5
6.00	6	37.5
Total	16	100.0

Survey D2T1

	Frequency	Percent
2.45	1	1.3
3.00	1	1.3
3.16	2	2.5
3.87	2	2.5
4.00	2	2.5
4.24	6	7.6
4.47	5	6.3
4.90	4	5.1
5.00	6	7.6
5.48	15	19.0
6.00	35	44.3
Total	79	100.0

Panel D2T2

	Frequency	Percent
3.46	2	12.5
3.87	1	6.3
4.00	1	6.3
4.24	1	6.3
4.47	1	6.3
4.90	2	12.5
5.00	1	6.3
5.48	3	18.8
6.00	4	25.0
Total	16	100.0

Survey D2T2

	Frequency	Percent
2.45	1	1.3
3.00	1	1.3
3.16	1	1.3
3.46	2	2.5
3.87	1	1.3
4.00	2	2.5
4.24	4	5.1
4.47	5	6.3
4.90	6	7.6
5.00	7	8.9
5.48	31	39.2
6.00	18	22.8
Total	79	100.0

Panel D2T3

	Frequency	Percent
2.45	1	6.3
2.83	1	6.3
3.46	2	12.5
3.87	5	31.3
4.00	1	6.3
4.47	2	12.5
5.00	2	12.5
5.48	1	6.3
6.00	1	6.3
Total	16	100.0

Survey D2T3

	Frequency	Percent
2.45	1	1.3
3.00	2	2.5
3.16	2	2.5
3.46	14	17.7
3.87	2	2.5
4.00	6	7.6
4.24	1	1.3
4.47	12	15.2
4.90	5	6.3
5.00	9	11.4
5.48	16	20.3
6.00	9	11.4
Total	79	100.0

Panel D2T4

	Frequency	Percent
4.47	2	12.5
4.90	4	25.0
5.48	3	18.8
6.00	7	43.8
Total	16	100.0

Survey D2T4

	Frequency	Percent
2.45	1	1.3
2.83	1	1.3
3.00	1	1.3
3.16	1	1.3
3.46	3	3.8
3.87	2	2.5
4.00	2	2.5
4.24	3	3.8
4.47	6	7.6
4.90	5	6.3
5.00	11	13.9
5.48	25	31.6
6.00	18	22.8
Total	79	100.0

Panel D2T5

	Frequency	Percent
3.16	1	6.3
3.46	2	12.5
3.87	1	6.3
4.24	2	12.5
4.47	3	18.8
5.00	1	6.3
5.48	4	25.0
6.00	2	12.5
Total	16	100.0

Survey D2T5

	Frequency	Percent
2.45	1	1.3
3.00	2	2.5
3.16	1	1.3
3.46	2	2.5
3.87	3	3.8
4.00	1	1.3
4.24	3	3.8
4.47	5	6.3
4.90	1	1.3
5.00	5	6.3
5.48	33	41.8
6.00	22	27.8
Total	79	100.0

Panel D2T6

	Frequency	Percent
3.16	1	6.3
3.87	1	6.3
4.00	1	6.3
4.47	3	18.8
4.90	1	6.3
5.00	1	6.3
5.48	6	37.5
6.00	2	12.5
Total	16	100.0

Survey D2T6

	Frequency	Percent
2.45	1	1.3
3.00	2	2.5
3.16	1	1.3
3.46	1	1.3
3.87	3	3.8
4.00	3	3.8
4.24	2	2.5
4.47	6	7.6
4.90	7	8.9
5.00	6	7.6
5.48	27	34.2
6.00	20	25.3
Total	79	100.0

Panel D3T1

	Frequency	Percent
4.90	2	12.5
5.00	2	12.5
5.48	6	37.5
6.00	6	37.5
Total	16	100.0

Survey D3T1

	Frequency	Percent
2.45	1	1.3
2.83	1	1.3
3.16	1	1.3
3.46	1	1.3
4.00	2	2.5
4.24	4	5.1
4.47	5	6.3
4.90	7	8.9
5.00	6	7.6
5.48	30	38.0
6.00	21	26.6
Total	79	100.0

Panel D3T2

	Frequency	Percent
3.87	1	6.3
4.47	1	6.3
5.00	1	6.3
5.48	5	31.3
6.00	8	50.0
Total	16	100.0

Survey D3T2

	Frequency	Percent
3.00	1	1.3
3.16	1	1.3
3.46	1	1.3
3.87	3	3.8
4.00	1	1.3
4.24	2	2.5
4.47	3	3.8
4.90	7	8.9
5.00	6	7.6
5.48	15	19.0
6.00	39	49.4
Total	79	100.0

Panel D3T3

	Frequency	Percent
3.87	1	6.3
4.00	1	6.3
4.24	4	25.0
5.48	5	31.3
6.00	5	31.3
Total	16	100.0

Survey D3T3

	Frequency	Percent
2.45	1	1.3
3.46	4	5.1
3.87	3	3.8
4.00	1	1.3
4.24	3	3.8
4.47	3	3.8
4.90	6	7.6
5.00	5	6.3
5.48	23	29.1
6.00	30	38.0
Total	79	100.0

Panel D3T4

	Frequency	Percent
3.87	1	6.3
4.24	1	6.3
4.47	1	6.3
4.90	2	12.5
5.00	3	18.8
5.48	5	31.3
6.00	3	18.8
Total	16	100.0

Survey D3T4

	Frequency	Percent
2.45	1	1.3
3.16	1	1.3
3.46	1	1.3
3.87	1	1.3
4.00	2	2.5
4.24	4	5.1
4.47	5	6.3
4.90	11	13.9
5.00	10	12.7
5.48	23	29.1
6.00	20	25.3
Total	79	100.0

Panel D3T5

	Frequency	Percent
3.46	2	12.5
3.87	1	6.3
4.00	1	6.3
4.24	1	6.3
4.90	1	6.3
5.00	2	12.5
5.48	1	6.3
6.00	7	43.8
Total	16	100.0

Survey D3T5

	Frequency	Percent
2.00	1	1.3
2.83	1	1.3
3.46	2	2.5
3.87	2	2.5
4.00	3	3.8
4.24	5	6.3
4.47	3	3.8
4.90	11	13.9
5.00	9	11.4
5.48	20	25.3
6.00	22	27.8
Total	79	100.0

Panel D3T6

	Frequency	Percent
3.00	1	6.3
3.46	1	6.3
3.87	1	6.3
4.00	3	18.8
4.90	2	12.5
5.48	5	31.3
6.00	3	18.8
Total	16	100.0

Survey D3T6

	Frequency	Percent
2.45	1	1.3
2.83	1	1.3
3.46	2	2.5
3.87	4	5.1
4.00	2	2.5
4.24	3	3.8
4.47	3	3.8
4.90	6	7.6
5.00	9	11.4
5.48	22	27.8
6.00	26	32.9
Total	79	100.0

Panel D3T7

	Frequency	Percent
3.46	1	6.3
3.87	3	18.8
4.47	1	6.3
5.00	1	6.3
5.48	5	31.3
6.00	5	31.3
Total	16	100.0

Survey D3T7

	Frequency	Percent
2.45	1	1.3
3.16	1	1.3
3.46	5	6.3
4.00	2	2.5
4.24	4	5.1
4.47	7	8.9
4.90	7	8.9
5.00	9	11.4
5.48	18	22.8
6.00	25	31.6
Total	79	100.0

Panel D4T1

	Frequency	Percent
4.24	2	12.5
4.47	2	12.5
5.00	1	6.3
5.48	2	12.5
6.00	9	56.3
Total	16	100.0

Survey D4T1

	Frequency	Percent
2.24	1	1.3
3.00	1	1.3
3.16	1	1.3
3.46	2	2.5
4.00	3	3.8
4.24	1	1.3
4.47	4	5.1
4.90	10	12.7
5.00	12	15.2
5.48	20	25.3
6.00	24	30.4
Total	79	100.0

Panel D4T2

	Frequency	Percent
3.87	1	6.3
4.00	1	6.3
4.24	1	6.3
4.47	1	6.3
4.90	2	12.5
5.48	2	12.5
6.00	8	50.0
Total	16	100.0

Survey D4T2

	Frequency	Percent
2.24	1	1.3
3.00	1	1.3
3.16	1	1.3
3.46	1	1.3
4.00	2	2.5
4.24	1	1.3
4.47	7	8.9
4.90	9	11.4
5.00	9	11.4
5.48	20	25.3
6.00	27	34.2
Total	79	100.0

Panel D5T1

	Frequency	Percent
3.46	2	12.5
4.00	5	31.3
4.47	2	12.5
4.90	1	6.3
5.48	5	31.3
6.00	1	6.3
Total	16	100.0

Survey D5T1

	Frequency	Percent
2.24	1	1.3
2.45	1	1.3
3.00	1	1.3
3.16	1	1.3
3.46	4	5.1
4.00	10	12.7
4.24	4	5.1
4.47	3	3.8
4.90	5	6.3
5.00	12	15.2
5.48	18	22.8
6.00	19	24.1
Total	79	100.0

Panel D5T2

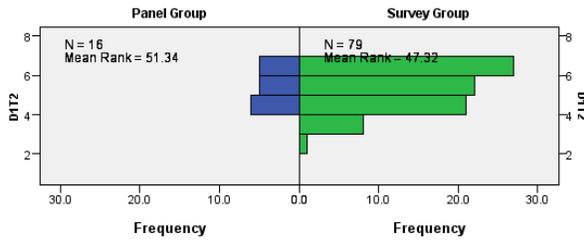
	Frequency	Percent
3.87	1	6.3
4.90	2	12.5
5.48	2	12.5
6.00	11	68.8
Total	16	100.0

Survey D5T2

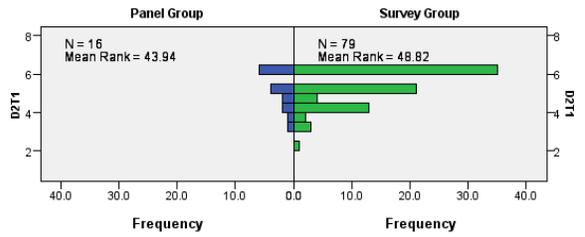
	Frequency	Percent
3.00	1	1.3
3.46	2	2.5
3.87	1	1.3
4.00	2	2.5
4.24	1	1.3
4.47	1	1.3
4.90	6	7.6
5.00	11	13.9
5.48	17	21.5
6.00	37	46.8
Total	79	100.0

APPENDIX S. GRAPHS COMPARING MEAN RANK IN MANN-WHITNEY U ANALYSIS

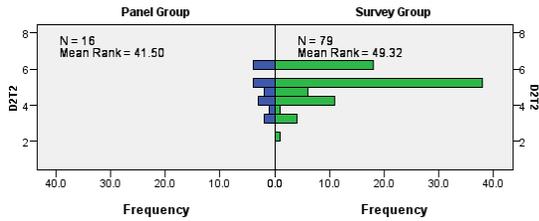
Independent-Samples Mann-Whitney U Test



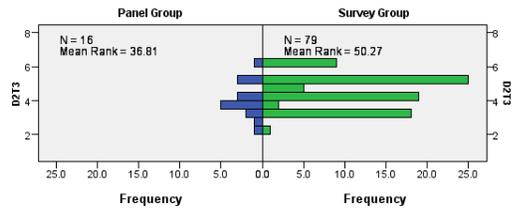
Independent-Samples Mann-Whitney U Test



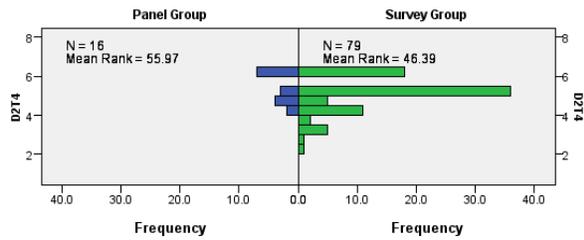
Independent-Samples Mann-Whitney U Test



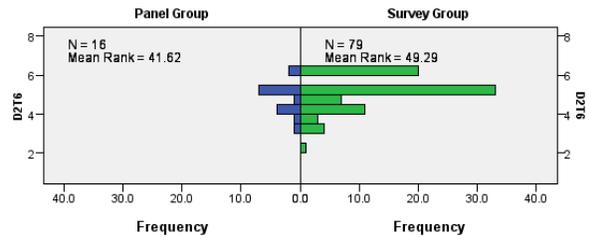
Independent-Samples Mann-Whitney U Test



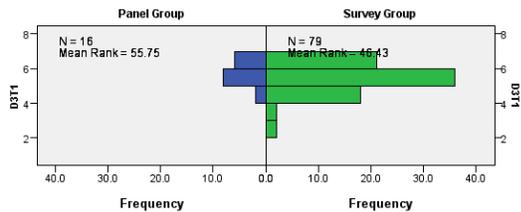
Independent-Samples Mann-Whitney U Test



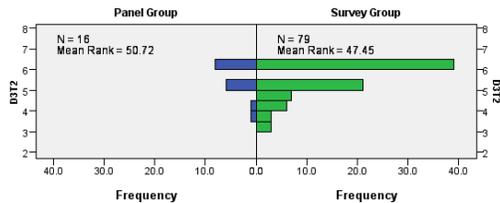
Independent-Samples Mann-Whitney U Test



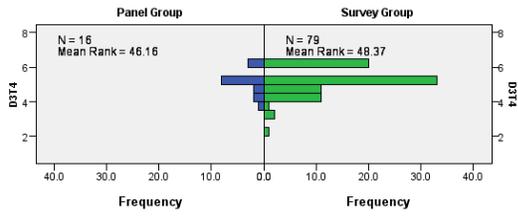
Independent-Samples Mann-Whitney U Test



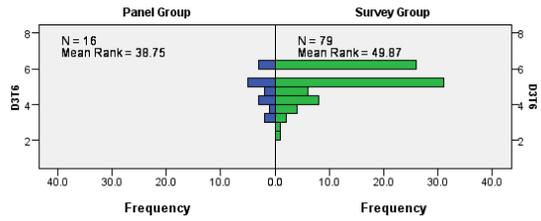
Independent-Samples Mann-Whitney U Test



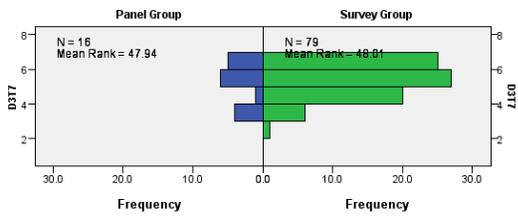
Independent-Samples Mann-Whitney U Test



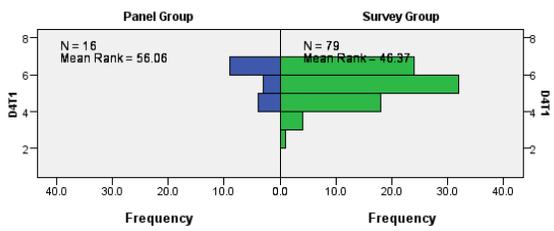
Independent-Samples Mann-Whitney U Test



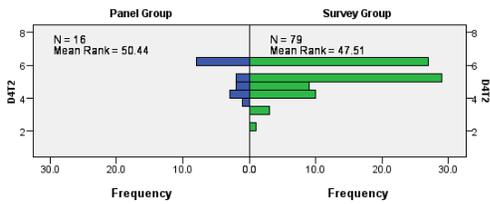
Independent-Samples Mann-Whitney U Test



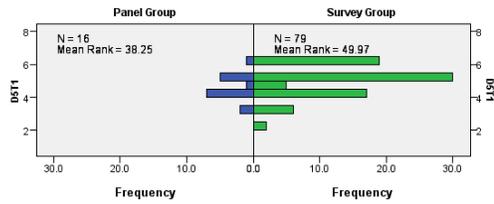
Independent-Samples Mann-Whitney U Test



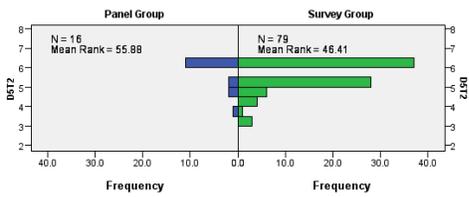
Independent-Samples Mann-Whitney U Test



Independent-Samples Mann-Whitney U Test



Independent-Samples Mann-Whitney U Test



APPENDIX T. ADDITIONAL SUGGESTIONS FROM SRVEY PARTICIPANTS FOR EACH CONTENT DOMAIN

D1Other

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid	72	91.1	91.1	91.1
Ability to demonstrate classroom/behavior management and see the "big picture".	1	1.3	1.3	92.4
Ability to observe children and determine when they are ready to move on to more challenging lessons is very important. If a teacher cannot competently do this then the child will suffer and the teacher will have a bored child which can in turn create behavior issues.	1	1.3	1.3	93.7
Being consistent is very important.	1	1.3	1.3	94.9
Competent mastery of materials is essential.				
Thorough understanding of the materials primary objective before introducing extensions.	1	1.3	1.3	96.2
No	1	1.3	1.3	97.5
The way a lesson is present is critical to the Montessori authenticity	1	1.3	1.3	98.7
When providing task that are developmentally appropriate, it allows the teacher to see where the child is and what they need help with.	1	1.3	1.3	100.0
Total	79	100.0	100.0	

D2Other

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid	76	96.2	96.2	96.2
Each classroom environment should be treated as a community. It should also be monitored and adjusted to fir the personalities of the students but at the same time keeping within the Montessori realms.	1	1.3	1.3	97.5
Reflection upon how many lessons are being presented and being sure all areas are presented to tap into the potenitality of each child's interest	1	1.3	1.3	98.7
The teacher should have the needed materials to set up the classroom properly. If she does not, her mentor should see to it that she gets those needed materials. The new teacher should be told to be vocal about these needs until they are met.	1	1.3	1.3	100.0
Total	79	100.0	100.0	

D3Other

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid	76	96.2	96.2	96.2
Being consistent.	1	1.3	1.3	97.5
Once classroom ground rules are established, rules are suppose to be followed. Every now and then they need to be refreshed, so that the students will know exactly what the rules are and exactly what they're suppose to do.	1	1.3	1.3	98.7
record keeping of who has jobs, wo has certain responsibilities and rotating them as needed	1	1.3	1.3	100.0
Total	79	100.0	100.0	

D4Other

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid	76	96.2	96.2	96.2
Ability to meet the demands of assessment required by public school systems	1	1.3	1.3	97.5
Monitoring all students is something that has to be done in a classroom. It doesn't matter whether it is a Montessori or a Traditional Classroom. You can serve your students better if you know them personally as well as academically!!	1	1.3	1.3	98.7
The mentor/team should make sure before school starts that the new teacher is aware of rules for CCDEP, etc..	1	1.3	1.3	100.0
Total	79	100.0	100.0	

D5Other

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid	77	97.5	97.5	97.5
Interact and communicate well with peers and administration	1	1.3	1.3	98.7
There should be a wonderful communication relationship between the student and the teacher. Students tend to do better when their teacher is someone they can talk to.	1	1.3	1.3	100.0
Total	79	100.0	100.0	

APPENDIX U. SUMMARY OF COMMENTS UNDER EACH DOMAIN IN THE SURVEY

Domain	Comment ID #	Other Task/Comment	Researcher Comment
1	1	When providing tasks that are developmentally appropriate, it allows the teacher to see where the child is and what they need help with.	Not a task, but a rationale for the existing task
	2	Being consistent is very important.	Not a task
	3	Competent mastery of materials is essential. Thorough understanding of the materials primary objective before introducing extensions.	Not a task, but knowledge and skills necessary to complete task
	4	Ability to demonstrate classroom/behavior management and see the "big picture".	Addressed in Domain 3
	5	Ability to observe children and determine when they are ready to move on to more challenging lessons is very important. If a teacher cannot competently do this then the child will suffer and the teacher will have a bored child which can in turn create behavior issues.	Addressed in Domain 4
	6	The way a lesson is present is critical to the Montessori authenticity	Not a tasks, but relates to knowledge and skills necessary to complete the task.
2	7	Each classroom environment should be treated as a community. It should also be monitored and adjusted to fir the personalities of the students but at the same time keeping within the Montessori realms.	A summary of what is incorporated in this domain and addressed in Domain 3
	8	The teacher should have the needed materials to set up the classroom properly. If she does not, her mentor should see to it that she gets those needed materials. The new teacher should be told to be vocal about these needs until they are met.	Not a task, but denotes the importance of a well-appointed classroom. Last sentence may relate to Domain 5 and communications.

	9	Reflection upon how many lessons are being presented and being sure all areas are presented to tap into the potentiality of each child's interest	Encompassed in the description of the domain and existing tasks.
3	10	Once classroom ground rules are established, rules are supposed to be followed. Every now and then they need to be refreshed, so that the students will know exactly what the rules are and exactly what they're supposed to do.	Encompassed in the description of the domain and existing tasks.
	11	Being consistent.	Not a task
	12	record keeping of who has jobs, who has certain responsibilities and rotating them as needed	This is part of the classroom community developed and maintained through the tasks in Domain 3. Record keeping related to individual students is also included in Domain 4.
4	13	Monitoring all students is something that has to be done in a classroom. It doesn't matter whether it is a Montessori or a Traditional Classroom. You can serve your students better if you know them personally as well as academically!!	General comment, not a task.
	14	The mentor/team should make sure before school starts that the new teacher is aware of rules for CCDEP, etc.	This is not a task for an entry-level teacher
	15	Ability to meet the demands of assessment required by public school systems	Panel made the decision not to include anything related to standardized testing in public schools as it is related to the specific role of an ECMT. This is not something all ECMTs deal with.

5	16	There should be a wonderful communication relationship between the student and the teacher. Students tend to do better when their teacher is someone they can talk to.	General teacher quality, not a specific task.
	17	Interact and communicate well with peers and administration	Task 5.2

REFERENCES

- AERA, APA, & NCME. (2014). Standards for educational and psychological testing. Washington, DC: American Educational Research Association.
- Althouse, L. A., Du, Y., & Ham, H. P. (2009). Confirming the validity of the general pediatrics certification examinations: A practice analysis. *The Journal of Pediatrics*, *155*(2), 155-156. doi:10.1016/j.jpeds.2009.04.028
- Andrich, D. (1988). A general form of Rasch's extended logistic model for partial credit scoring. *Applied Measurement in Education*, *1*(4), 363.
- Arbet, S., Lathrop, J., & Hooker, R. S. (2009). Using practice analysis to improve the certifying examinations for PAs. *Journal of the American Association of Physician Assistants*, *22*(2), 31-36.
- Armstead V. Starkville Municipal Separate School District, No. 325 F. Supp. 560 (U.S. District Court for the Northern District of Mississippi 1971).
- Babcock, B., Koenen, R., Vatterott, T. L., Belcher, T. G., Current, L., Merck, G. A., . . . Wartenbee, S. R. (2013). Summary of ARRT's 2012 practice analysis for bone densitometry. *Radiologic Technology*, *84*(4), 431-433.
- Babcock, B., & Yoes, M. E. (2013). Enhancing job analysis surveys in the medical specialties with CMS data. *Evaluation & the Health Professions*, *36*(2), 240-246. doi:10.1177/0163278712446964
- Barnsteiner, J. H., & Wyatt, J. S. (2002). What do pediatric nurses do? Results of the role delineation study in Canada and the United States. *Credentialing and Professionalism in Pediatric Nursing*, *28*(2), 165-170.
- Battenfield, A. E., & Schehl, J. (2013). Practice analysis for human performance technologists. *Performance Improvement*, *52*(8), 15-20. doi:10.1002/pfi.21366
- Bevans, M., Hastings, C., Wehrlen, L., Cusack, G., Matlock, A. M., Miller-Davis, C., . . . Wallen, G. R. (2011). Defining Clinical Research Nursing Practice: Results of a Role Delineation Study. *CTS: Clinical & Translational Science*, *4*(6), 421-427. doi:10.1111/j.1752-8062.2011.00365.x
- Bond, T. T., & Fox, C. M. (2007). *Applying the Rasch model: Fundamental measurement in the human sciences* (2nd ed.). Mahwah, New Jersey: Lawrence Erlbaum Associates.
- Borsboom, D., Mellenbergh, G. J., & van Heerden, J. (2004). The concept of validity. *Psychological Review*, *111*(4), 1061-1071. doi:10.1037/0033-295X.111.4.1061
- Bradley, P. D., Bergen, L. P., Ginter, E. J., Williams, L. M., & Scalise, J. J. (2010). A survey of North American marriage and family therapy practitioners: A role delineation study.

- American Journal of Family Therapy*, 38(4), 281-291.
doi:10.1080/01926187.2010.493119
- Brennan, R. L. (2013). Commentary on "Validating the interpretations and uses of test scores". *Journal of Educational Measurement*, 50(1), 74-83.
- Brown, C. G., Cantril, C., McMullen, L., Barkley, D. L., Dietz, M., Murphy, C. M., & Fabrey, L. J. (2012). Oncology nurse navigator role delineation study. *Clinical Journal of Oncology Nursing*, 16(6), 581-585.
- Buckendahl, C. W., & Hunt, R. (2005). Whose rules? The relation between the "rules" and "law" of testing. In R. P. Phelps (Ed.), *Defending standardized testing* (pp. 147-158). Mahwah, NJ: Lawrence Erlbaum Associates, Inc. .
- Buckendahl, C. W., & Plake, B. S. (2006). Evaluating tests. In S. M. Downing & T. M. Haladyna (Eds.), *Handbook of test development* (pp. 725-738). Mahwah, New Jersey: Lawrence Erlbaum Associates.
- Castro, K., Bevans, M., Miller-Davis, C., Cusack, G., Loscalzo, F., Matlock, A. M., . . . Hastings, C. (2011). Validating the clinical research nursing domain of practice. *Oncology Nursing Forum*, 38(2), E72-E80. doi:10.1188/11.ONF.E72-E80
- Civil Rights Act of 1964, 7, 42, U.S.C. 2000e et seq Stat. (1964).
- Cizek, G. J. (2001). Conjectures on the rise and fall of standard setting: An introduction to context and practice. In G. J. Cizek (Ed.), *Setting performance standards: Theory and applications*. London: Routledge.
- Cizek, G. J. (2012). Defining and distinguishing validity: Interpretations of score meaning and justifications of test use. *Psychological Methods*, 17(1), 31-43.
- Clauser, B. E., Margolis, M. J., & Case, S. M. (2006). Testing for licensure and certification in the professions. In R. L. Brennan (Ed.), *Educational Measurement* (4th ed.). Westport, CT: Praeger.
- Colton, A., Kane, M., Kingsbury, C., & Estes, C. A. (1991). A strategy for examining the validity of job analysis data. *Journal of Educational Measurement*, 28(4), 283-294.
doi:10.2307/1434893
- Cook, C., Heath, F., Thompson, R. L., & Thompson, B. (2001). Score Reliability in Webor Internet-Based Surveys: Unnumbered Graphic Rating Scales versus Likert-Type Scales. *Educational and Psychological Measurement*, 61(4), 697-706.
doi:10.1177/00131640121971356
- Corder, G. W., & Foreman, D. I. (2014). *Nonparametric statistics: A step-by-step approach* (2nd ed.). New Jersey: Wiley.

- Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, 52, 281-302.
- Darling, C. A., Fleming, W. M., & Cassidy, D. (2009). Professionalization of family life education: Defining the field. *Interdisciplinary Journal of Applied Family Studies*, 58, 330-345.
- Educational Testing Services. (2015). Praxis subject assessments test content and structure. Retrieved from www.ets.org/praxis/about/subject/content?WT.ac=31040_praxis_about_praxisii_content
- Equal Employment Opportunity Commission. (August 25, 1978). *Uniform guidelines on employee selection procedures*. 29 CFR part 1607.
- Erickson, H. L., Erickson, M. E., Campbell, J. A., Brekke, M. E., & Sandor, M. K. (2013). Validation of holistic nursing competencies: role-delineation study, 2012. *Journal of Holistic Nursing*. doi:10.1177/0898010113491144
- Fein, M. (2012). *Test development: Fundamentals for certification and evaluation*. Danvers, MA: American Society for Training and Development.
- Field, A., Miles, J., & Field, Z. (2012). *Discovering Statistics Using R*. Los Angeles: Sage.
- Flanagan, J. C. (1954). The critical incident technique. *Psychological Bulletin*, 51(4), 327-358. doi:10.1037/h0061470
- Gael, S. (1983). *Job analysis : A guide to assessing work activities*. San Francisco: Jossey-Bass.
- Gael, S. (1988a). Subject matter expert conferences. In S. Gael (Ed.), *The job analysis handbook for business, industry, and government* (Vol. 1, pp. 432-445). White Plains, NY: Wiley & Sons, Inc.
- Gael, S. (Ed.) (1988b). *The job analysis handbook for business, industry, and government* (Vol. 1). White Plains, NY: Wiley & Sons, Inc.
- Gael, S. (Ed.) (1988c). *The job analysis handbook for business, industry, and government* (Vol. 2). White Plains, NY: Wiley & Sons, Inc.
- Garbin, M. G., & Chmielewski, C. M. (2013). Job analysis and role delineation: LPN/LVNs and hemodialysis technicians. *Nephrology Nursing Journal*, 40(3), 225-240.
- Gerrow, J. D., Murphy, H. J., & Boyd, M. A. (2006). Competencies for the beginning dental practitioner in Canada: A validity survey. *Journal of Dental Education*, 70(10), 1076-1080.
- Gorman, S. L., Wruble Hakim, E., Johnson, W., Bose, S., Harris, K. S., Crist, M. H., . . . Coe, J. B. (2010). Nationwide acute care physical therapist practice analysis identifies

- knowledge, skills, and behaviors that reflect acute care practice. *Physical Therapy*, 90(10), 1453-1467. doi:10.2522/ptj.20090385
- Guardians v. Civil Service Commission of the City of New York, No. 630 F.2d 79 (U.S. Cour of Appeals, Second Circuit 1980).
- Gulino et al. v. Board of Education of the City School District of the City of New York, No. 2015 WL 3536694 (U.S. District Court, S.D. New York 2015).
- Harwell, M. R., & Gatti, G. G. (2001). Rescaling ordinal data to interval data in educational research. *Review of Educational Research*, 71(1), 105-131.
- Hawkins-Walsh, E., & Van Cleve, S. (2013). The pediatric mental health specialist: Role delineation. *The Journal for Nurse Practitioners*, 9(3), 142-148. doi:10.1016/j.nurpra.2012.11.017
- Henman, L. J., Corrigan, R., Carrico, R., & Suh, K. N. (2015). Identifying changes in the role of the infection preventionist through the 2014 practice analysis study conducted by the Certification Board of Infection Control and Epidemiology, Inc. *American Journal of Infection Control*, 43(7), 664-668. doi:10.1016/j.ajic.2015.02.026
- Hockberger, R. S., La Duca, A., Orr, N. A., Reinhart, M. A., & Sklar, D. P. (2003). Creating the model of a clinical practice: The case of emergency medicine. *Academic Emergency Medicine*, 10(2), 161-168. doi:10.1197/aemj.10.2.161
- Howell, D. C. (2002). *Statistical methods for psychology* (5th ed.). Pacific Grove, CA: Duxbury.
- Hsu, C., & Sandford, B. (2007). The delphi technique: Making sense of consensus. *Practice Assessment, research, & evaluation*, 12(10), 1-8.
- IBM Corp. (2013). IBM SPSS statistics for Windows, version 22.0. Armonk, NY: IBM Corp.
- International Standard Organization Committee on Conformity Assessment. (2012). Conformity assessment: General requirements for bodies operating certification of persons *ISO/IEC 17024*. Geneva, Switzerland: International Standards Organization.
- Johanson, M. A., Miller, M. B., Coe, J. B., & Campo, M. (2016). Orthopaedic physical therapy: Update to the description of specialty practice. *Journal of Orthopaedic & Sports Physical Therapy*, 46(1), 9-18. doi:doi:10.2519/jospt.2016.6211
- Johnson, S. B. (2008). *Workshop feedback questionnaire*. Unpublished Instrument.
- Jones, P., Smith, R. W., & Talley, D. (2006). Developing test forms for small-scale achievement testing systems. In S. M. Downing & T. M. Haladyna (Eds.), *Handbook of test development* (pp. 487-525). Mahwah, New Jersey: Lawrence Erlbaum Associates.
- Kane, M. (1992). An argument-based approach to validity. *Psychological Bulletin*, 112(3), 527-535. doi:10.1037/0033-2909.112.3.527

- Kane, M. (1994a). Validating interpretive arguments for licensure and certification examinations. *Evaluation & the Health Professions, 17*(2), 133-159. doi:10.1177/016327879401700202
- Kane, M. (1994b). Validating the performance standards associated with passing scores. *Review of Educational Research, 64*(3), 425-461. doi:10.2307/1170678
- Kane, M. (1997). Model-based practice analysis and test specifications. *Applied Measurement in Education, 10*(1), 5.
- Kane, M. (2006). Validation. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp. 17-64). Westport, Ct: Praeger.
- Kane, M. (2009). Validating the interpretations and uses of test scores. In R. W. Lissitz (Ed.), *The concept of validity* (pp. 39-64). Charlotte, NC: Information Age.
- Kane, M. (2013). Validating the interpretations and uses of test scores. *Journal of Educational Measurement, 50*(1), 1-73.
- Keller, J. L., Caro, C. M., Dimick, M. P., Landrieu, K., Fullenwider, L., & Walsh, J. M. (2016). Thirty years of hand therapy: The 2014 practice analysis. *Journal of Hand Therapy, 29*(3), 222-234. doi:10.1016/j.jht.2016.02.011
- Kenward, K. (2007). *Role delineation study of nurse practitioners and clinical nurse specialists*. Retrieved from Chicago, IL: https://ncsbn.org/06_LPN_RoleDelStudy_NCLEX_30_Web.pdf
- Knapp, J. E., & Knapp, L. G. (1995). Practice analysis: Building the foundation for validity. In J. C. Impara (Ed.), *Licensure testing purposes, procedures, and practices* (pp. 93-116). Lincoln, NE: Buros.
- Koby, G. S., & Melby, A. K. (2013). Certification and job task analysis (JTA): Establishing validity of translator certification examinations. *The International Journal for Translation & Interpreting Research, 5*(1), 174-210. doi:ti.105201.2013.a10
- Kramer, R. (1976). *Maria Montessori: A biography*. Chicago, IL: University of Chicago Press.
- LaDuca, T. (1994). Validation of professional licensure examinations. *Evaluation & the Health Professions, 17*(2), 178-197.
- LaDuca, T. (2006). Commentary: A closer look at task analysis: Reactions to Wang, Schnipke, and Witt. *Educational Measurement: Issues and Practice, 25*(2), 31-33. doi:10.1111/j.1745-3992.2006.00055.x
- Lane, S., Raymond, M. R., & Haladyna, T. M. (2016). *Handbook of test development* (2nd ed.). New York, NY: Routledge.

- Li, X., & Sireci, S. G. (2013). A new method for analyzing content validity data using multidimensional scaling. *Educational & Psychological Measurement, 73*(3), 365-385. doi:10.1177/0013164412473825
- Lillard, A. (2005). *Montessori: The science behind the genius*. New York, NY: Oxford University Press.
- Linacre, J. M. (2016a). WinstepsR (Version 3.92.0) [Computer Software]. Beaverton, OR: Winsteps.com.
- Linacre, J. M. (2016b). Winsteps^R Rasch measurement computer program user's guide. Beaverton, Oregon: Winsteps.com.
- Lissitz, R. W., & Samuelsen, K. (2007). A Suggested Change in Terminology and Emphasis Regarding Validity and Education. *Educational Researcher, 36*(8), 437-448. doi:10.3102/0013189x07311286
- Lund, A., & Lund, M. (2013). Mann-Whitney U Test. *Laerd Statistics*. Retrieved from <https://statistics.laerd.com/premium/spss/mwut/mann-whitney-test-in-spss-13.php>
- Lunz, M. E., Stahl, J. A., & James, K. (1989). Content validity revisited: Transforming job analysis data into test specifications. *Evaluation & the Health Professions, 12*(2), 192-206. doi:10.1177/016327878901200205
- MACTE. (2013). Self study checklist. Charlottesville, VA: Montessori Accreditation Council for Teacher Education.
- MACTE. (2015a). Retrieved from <http://www.macte.org/congratulations-to-the-state-of-wi-and-wi-montessori-alliance/>
- MACTE. (2015b). A policymaker's guide to Montessori teacher recognition: Becoming policy inclusive in early childhood education workforce registries & career pathways. Retrieved from <http://www.macte.org/release-of-new-document-a-policy-makers-guide-to-montessori-teacher-recognition/>
- Matarese, L. E., Chinn, R. N., Hertz, N. R., Callahan, P., Harvey-Banchik, L., & Strang, B. (2012). Practice analysis of nutrition support professionals: Evidence-based multidisciplinary nutrition support certification examination. *Journal of Parenteral and Enteral Nutrition, 36*(6), 663-670. doi:10.1177/0148607111435330
- Maurer, T., & Tross, S. (2000). SME committee vs. field job analysis ratings: Convergence, cautions, and a call. *Journal of Business and Psychology, 14*(3), 489-499. doi:10.1023/A:1022936418249
- McKenzie, J. F., Dennis, D., Auld, M. E., & Lysoby, L. (2016). Health education specialist practice analysis 2015 (HESPA 2015): Process and outcomes. *Health Education & Behavior, 43*(3), 286-295. doi:10.1177/1090198116639258

- Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (Vol. 3rd, pp. 13). New York, NY: Macmillan.
- Montessori, M. (1948). *Discovery of the child*. New York: Ballantine Books.
- Muckle, T. J., Apatov, N. M., & Plaus, K. (2009). A Report on the CCNA 2007 professional practice analysis. *AANA Journal*, pp. 181-189.
- Muckle, T. J., Plaus, K. A., Henderson, J., & Waters, E. (2012). Professional practice analysis: Determining job relatedness of the certification examination for nurse anesthetists. *Journal of Nursing Regulation*, 3(3), 55-61. doi:10.1016/S2155-8256(15)30199-X
- Muenzen, P. M., Corrigan, M. M., Smith, M. A. M., & Rodrigue, P. G. (2005). Updating the pharmacy technician certification examination: A practice analysis study. *American Journal of Health-System Pharmacy*, 62(23), 2542. doi:10.2146/ajhp050434
- National Center for Montessori in the Public Sector. (2014). 2014 Census Data Snapshot. Retrieved from <http://www.public-montessori.org/public-montessori-census-snapshot-2014>
- National Commission for Certifying Agencies. (2004). Standards for the accreditation of certification programs. Washington, DC: Institute for Credentialing Excellence.
- Nelson, D. S. (1994). Job analysis for licensure and certification exams: Science or politics? *Educational Measurement: Issues and Practice*, 13(3), 29-35.
- Ortelli, T. A. (2006). Defining the professional responsibilities of academic nurse educators: The results of a national practice analysis. *Nursing Education Perspectives*, 27(5), 242-246.
- Polk Lillard, P. (1996). *Montessori today*. New York, NY: Schoken Books.
- Qualtrics. (2009). Qualtrics Software. Provo, UT: Qualtrics Labs, Inc.
- Rambush, N. M. (1962). *Learning how to learn: An American approach to Montessori*. Baltimore, MD: Helicon Press.
- Raymond, M. R. (2001). Job analysis and the specification of content for licensure and certification examinations. *Applied Measurement in Education*, 14(4), 369.
- Raymond, M. R. (2002). A practical guide to practice analysis for credentialing examinations. *Educational Measurement: Issues and Practice*, 21(3), 25-37.
- Raymond, M. R. (2005). An NCME instructional module on developing and administering practice analysis questionnaires. *Educational Measurement: Issues and Practice*, 24(2), 29-42. doi:10.1111/j.1745-3992.2005.00009.x

- Raymond, M. R. (2016). Job analysis, practice analysis, and the content in credentialing examinations. In S. Lane, M. R. Raymond, & T. M. Haladyna (Eds.), *Handbook of Test Development* (pp. 144-164). New York, NY: Routledge.
- Raymond, M. R., & Neustal, S. (2006). Determining the content fo credentialing examinations. In S. M. Downing & T. M. Haladyna (Eds.), *Handbook of test development* (pp. 181-223). Mahwah, New Jersey: Lawrence Erlbaum Associates.
- Reuter-Rice, K., Madden, M. A., Gutknecht, S., & Foerster, A. (2016). Acute care pediatric nurse practitioner: The 2014 practice analysis. *Journal of Pediatric Health Care, 30*(3), 241-251. doi:10.1016/j.pedhc.2016.01.009
- Ricci v. Destafano et al., 557, (U.S. Supreme Court, 2009).
- Schafer, W. D., Wang, J., & Wang, V. (2009). Validity in action: State assessment validity evidence for compliance with NCLB. In R. W. Lissitz (Ed.), *The concept of validity: Revisions, new directions, and applications* (pp. 173-194). Charlotte, NC: Information Age Publishing.
- Shreiner, B., Kolb, L., O'Brian, C. A., Carroll, S., & Lipman, R. D. (2015). National role delineation study of the board certification for advanced diabetes management. *The Diabetes Educator, 41*(5), 609-615. doi:10.1177/0145721715599269
- Sireci, S. G. (2007). On validity theory and test validation. *Educational Researcher, 36*(8), 477-481.
- Sireci, S. G., & Parker, P. (2006). Validity on trial: Psychometric and legal conceptualizations of validity. *Educational Measurement: Issues and Practice, 25*(3), 27-34. doi:10.1111/j.1745-3992.2006.00065.x
- South Carolina Montessori Alliance. (2015). SCMA Schools. Retrieved from <http://www.scmontessori.org/scma-schools/>
- Spray, J. A., & Huang, C.-Y. (2000). Obtaining test blueprint weights from job analysis surveys. *Journal of Educational Measurement, 37*(3), 187-201. doi:10.2307/1435287
- Starling, L. A., Waliczek, T. M., Haller, R., Brown, B. J., Malone, R., & Mitrione, S. (2014). Job task analysis survey for the horticultural therapy profession. *Horttechnology, 24*(6), 645-654.
- State of South Carolina Government. (2016). Retrieved from <http://www.sc.gov/government/Local/Pages/default.aspx>
- Tahan, H. M., Watson, A. C., & Sminkey, P. V. (2015). What case managers should know about their role and functions: A national study from the commission for case manager certification Part 2. *Professional Case Management, 20*(6), 271-296. doi:10.1097/NCM.000000000000115

- Tahan, H. M., Watson, A. C., & Sminkey, P. V. (2016). Informing the Content and Composition of the CCM Certification Examination: A National Study From the Commission for Case Manager Certification: Part 2. *Professional Case Management*, 21(1), 3-21. doi:10.1097/ncm.0000000000000129
- Tannenbaum, R. J., & Wesley, S. (1993). Agreement between committee-based and field-based job analyses: A study in the context of licensure testing. *Journal of Applied Psychology*, 78(6), 975-980. doi:10.1037/0021-9010.78.6.975
- Taub, A., Gilmore, G. D., Olsen, L. K., & Connell, D. (2015). Methodological comparisons of four national health education specialist role delineation research projects (1978-2012) in the United States. *Pedagogy in Health Promotion*, 1(4), 173-185. doi:10.1177/2373379915611681
- Taub, A., Olsen, L. K., Gilmore, G. D., & Connell, D. (2008). A role delineation study of health educators in the United States: methodological innovations. *Promotion & Education*, 15(4), 38-43. doi:10.1177/1025382308097697
- The Riley Institute. (2014). *The South Carolina public Montessori study: 2013-2014 teacher survey*. Retrieved from https://riley.furman.edu/sites/default/files/docs/MontTeacherSurveySummary13-14_ADS%2012%2017.pdf
- Tony, C. M. L., & Klockars, A. J. (1982). Anchor point effects on the equivalence of questionnaire items. *Journal of Educational Measurement*, 19(4), 317-322.
- Tsai, T.-H., & Kramer, G. A. (2014). Content validity of the national board dental hygiene examination. *Evaluation & the Health Professions*, 37(4), 503-523. doi:10.1177/0163278713501073
- United States Census Bureau. (2016). QuickFacts. <http://www.census.gov/quickfacts/table/PST045215/4525810,4549570,45065,00>
- Universal Public Purchasing Certification Council Governance Board. (2009). Job analysis for the CPPB and CPPO examinations. *Journal of Public Procurement*, 9(3 & 4), 289-325.
- Vandaveer, V. V., Lowman, R. L., Pearlman, K., & Brannick, J. P. (2016). A practice analysis of coaching psychology: Toward a foundational competency model. *Consulting Psychology Journal: Practice and Research*, 68(2), 118-142. doi:10.1037/cpb0000057
- Villanueva, N. E., Thompson, H. J., Macpherson, B. C., Meunier, K. E., & Hilton, E. (2006). The neuroscience nursing 2005 role delineation study: Implications for certification. *Journal of Neuroscience Nursing*, 38(6), 403-408, 415.
- Wang, N., Schnipke, D., & Witt, E. A. (2005). Use of knowledge, skill, and ability statements in developing licensure and certification examinations. *Educational Measurement: Issues and Practice*, 24(1), 15-22. doi:10.1111/j.1745-3992.2005.00003.x

- Wang, N., & Stahl, J. (2012). Obtaining content weights for test specifications from job analysis task surveys: An application of the many-facets Rasch model. *International Journal of Testing, 12*(4), 299-320. doi:10.1080/15305058.2011.639472
- Wang, N., Witt, E. A., & Schnipke, D. (2006). Rejoinder: A further discussion of job analysis and use of ksas in developing licensure and certification examinations: A response to LaDuca. *Educational Measurement: Issues and Practice, 25*(2), 34-37. doi:10.1111/j.1745-3992.2006.00056.x
- Whitescarver, K., & Cossentino, J. (2008). Montessori and the mainstream: A century of reform on the margins. *Teachers College Record, 110*(12), 2571-2600.
- Willens, J. S., DePascale, C., & Penny, J. (2010). Role delineation study for the American society for pain management nursing. *Pain Management Nursing, 11*(2), 68-75. doi:10.1016/j.pmn.2010.03.010
- Wolever, R. Q., Jordan, M., Lawson, K., & Moore, M. (2016). Advancing a new evidence-based professional in health care: Job task analysis for health and wellness coaches. *BMC Health Services Research, 16*(205), 1-11. doi:10.1186/s12913-016-1465-8
- Wright, B., & Panchapakesan, N. (1969). A procedure for sample-free item analysis. *Educational & Psychological Measurement, 29*(1), 23-48.
- Yang, J. C., Kazerooni, E. A., Bosma, J. L., Gerdeman, A. M., Becker, G. J., & Vydareny, K. H. (2012). Practice analysis: A basis for content validity for American Board of Radiology examinations in diagnostic radiology. *Journal of the American College of Radiology, 9*(2), 121-128. doi:10.1016/j.jacr.2011.08.020
- Zrebiec, J. (2014). A national study of the certified diabetes educator: Implications for future certification examinations. *The Diabetes Educator*. doi:10.1177/0145721714534783