

January 2013

# Preparing Teachers to Apply Research to Mathematics Teaching: Using Design-Based Research to Define and Assess the Process of Evidence-Based Practice

Sarah Van Ingen

*University of South Florida*, [vaningen@usf.edu](mailto:vaningen@usf.edu)

Follow this and additional works at: <http://scholarcommons.usf.edu/etd>

 Part of the [Science and Mathematics Education Commons](#), and the [Teacher Education and Professional Development Commons](#)

---

## Scholar Commons Citation

Van Ingen, Sarah, "Preparing Teachers to Apply Research to Mathematics Teaching: Using Design-Based Research to Define and Assess the Process of Evidence-Based Practice" (2013). *Graduate Theses and Dissertations*.  
<http://scholarcommons.usf.edu/etd/4799>

This Dissertation is brought to you for free and open access by the Graduate School at Scholar Commons. It has been accepted for inclusion in Graduate Theses and Dissertations by an authorized administrator of Scholar Commons. For more information, please contact [scholarcommons@usf.edu](mailto:scholarcommons@usf.edu).

Preparing Teachers to Apply Research to Mathematics Teaching: Using Design-Based  
Research to Define and Assess the Process of Evidence-Based Practice

by

Sarah van Ingen

A dissertation submitted in partial fulfillment  
of the requirements for the degree of  
Doctor of Philosophy  
Department of Secondary Education  
College of Education  
University of South Florida

Major Professor: Gladis Kersaint, Ph.D.  
Eugenia Vomvoridi- Ivanovic, Ph.D.  
Jeffrey Kromrey, Ph.D.  
Denisse R.Thompson, Ph.D.

Date of Approval:  
June 3, 2013

Keywords: Generalizability Theory, Mathematics Education, Performance Assessment,  
Research Utilization, Routines of Practice, Teacher Education

Copyright © 2013, Sarah van Ingen

## **Dedication**

To my husband, Daniel, my forever partner on this journey— I love you with all my heart and I could not have done this without your encouragement, steadfast love, and great sense of humor. You were there for me through it all! Thank you for all of the sacrifices and for the joy of sharing life together.

To my children—you are precious beyond measure. I am so grateful for your laughter and for the incredible gift it is to spend time with you. Being with you brings so much joy to my life.

To my first teacher Mary Sue Dobbin— it is only now that I have begun to understand the courage you had and the gift you gave me. You didn't see your own blindness as a limitation and you taught me both to read and to celebrate challenge and possibility.

And to classroom mathematics teachers who courageously fight for the best learning opportunities for their students—this document flows from my commitment to support and honor your work.

## **Acknowledgments**

There are times when words cannot adequately express the gratitude in one's heart, and this is one of them! There are many educators whom I wish to thank: First to my major professor Dr. Gladis Kersaint, I am so grateful for your expert guidance and the freedom you gave me to pursue this research agenda; to my committee members Drs. Eugenia Vomvoridi-Ivanovic, Jeff Kromrey, and Denisse Thompson, this document is much stronger because of your wisdom and support; to Susan Ariew, I am so grateful for our energizing meetings and the opportunity to work with and learn from you; to all of my USF professors, including Drs. Patty McHatton, Howard Johnston, and Robert Dedrick; to Dr. Danielle Parrish, thank you for your help and inspiring work; to the many educators including Dr. Ted and Michelle Johnson, Penny Ardnt, and Mary Faith Hall who prepared me for graduate studies. I am so grateful for the help of fellow doc students, especially Amanda Loyden, Elaine Cerrato Fisher, George McDonald and Sarah Bleiler—Sarah you have been the truest of friends, I would not have made it without you, and I continue to be inspired by both your intellect and your kindness. I am so grateful for the support and love of my parents, my siblings, the encouragement of my parents-in-law Dr. John and Maxine van Ingen, Susan and Bill Rademacher—prophetically you said that my time would come, and my dearest friend Christy Hamlin—you had my back and helped me learn how to run this race with great joy. Finally, the deepest gratitude in my heart goes to the One in whom I live and move and have my being. On this step of the journey I have learned to see a little more clearly that Love never fails.

## Table of Contents

List of Tables .....	iv
List of Figures .....	vi
Abstract .....	vii
Chapter 1: Introduction .....	1
Problem Statement: Persistent Disparities in Mathematics Achievement .....	1
The Challenge of Linking Research to Teaching Practice .....	5
Previous Approaches to Linking Research to Practice .....	11
The Process of Evidence-Based Practice .....	12
Conceptual Framework .....	14
Research Purpose and Study Design .....	16
Research Questions .....	18
Delimitations and Study Significance .....	18
Limitations .....	20
Chapter 2: Literature Review .....	21
Preparing Practitioners for Evidence-Based Practice .....	22
Defining evidence-based practice .....	22
Outcomes of training on EBP .....	24
Clinical outcomes .....	24
Educational outcomes .....	26
Systematic reviews .....	26
Individual studies .....	28
Theoretical and contextual considerations for training on EBP .....	29
Challenges to EBP in social work .....	30
Social work field placements and training on EBP .....	32
Linking Research to Practice in Teaching and Teacher Preparation .....	33
Linking research to practice with inservice teachers .....	34
How do inservice teachers understand research? .....	34
How do inservice teachers implement education research? .....	37
What are the social contexts in which inservice teachers link research to practice? .....	39
Top-down leadership approaches to research implementation .....	39
Professional relationships that influence research implementation .....	41
Systematic change .....	44

Linking research to practice with preservice teachers .....	45
The meaning of education research .....	45
Experiential learning about applying research to practice .....	46
Longitudinal studies.....	50
Routines of Practice .....	51
Performance Assessments in Teacher Preparation .....	56
Defining performance assessments.....	56
Validity and performance assessments .....	58
Reliability and performance assessments .....	61
Performance assessments in the context of TPPs .....	65
Initial adoption of a performance assessment.....	65
Validity in TPP performance assessments .....	68
Reliability in TPP performance assessments .....	70
Design-Based Research .....	72
Terminology and origins.....	73
Elements of DBR .....	73
Challenges and indications of quality.....	77
Discussion .....	79
Preparing practitioners for EBP .....	79
Linking research to practice in teacher preparation.....	81
Routines of practice .....	83
Performance assessments in teacher preparation.....	85
Design-based research .....	86
 Chapter 3: Method .....	 89
The First Four DBR Iterations .....	92
Iteration 1 .....	94
Design of the intervention.....	94
Ecological environment .....	95
Avenues of inquiry.....	96
Quasi-experimental study .....	96
Grounded theory analysis if ERP research application	
essays .....	101
Treatment potency test.....	103
Iteration 2 .....	104
Iteration 3 .....	105
Iteration 4 .....	109
Iteration 5 .....	112
Participants.....	112
Intervention .....	112
Ecological environment .....	114
System of assessment.....	115
Instrumentation .....	117
Data collection .....	119
Data analysis .....	119
The teaching intervention .....	120

The system of assessment .....	122
Summary .....	125
Chapter 4: Results .....	127
The Information Literacy Questionnaire- Short Version.....	127
Descriptive statistics .....	127
Inferential statistics .....	134
Familiarity with the Process of Evidence-Based Practice in Education Scale ....	135
Education Research Project- Residency Version (ERP-RV).....	136
Standardized Performance Assessment .....	142
Chapter 5: Discussion .....	148
Interpretation of Results.....	148
Question 1 .....	149
Question 2 .....	154
Question 3 .....	156
Question 4 .....	158
Research Implications .....	161
Preparing practitioners for evidence-based practice .....	162
Linking research to practice in teacher preparation .....	163
Routines of practice .....	165
Performance assessments.....	166
Design-based research .....	167
Practice Implications.....	168
Directions for Future Research .....	170
Conclusion .....	172
References.....	174
Appendices.....	196
Appendix A: Education Research Project (Iteration 1) .....	197
Appendix B: Information Literacy Questionnaire Pilot Version (Iteration 1).....	203
Appendix C: Original Rubric for Evaluation of First Four Tasks from Education Research Project (Iteration 1) .....	207
Appendix D: Codebook for Grounded Theory Study (Iteration 1).....	208
Appendix E: Relationships among Codes, Categories and Themes from Grounded Theory Analysis of Research Application Essays (Iteration 1).....	212
Appendix F: Information Literacy Questionnaire Short Version (Iterations 2-5) .....	213
Appendix G: Interview Protocol for Students Who Completed Education Research Project.....	215
Appendix H: Rubric for Formative Assessment of the Revised Education Research Project (Iteration 4) .....	216
Appendix I: Research Summaries for Summative Assessment of the Process of Evidence-Based Practice .....	218

Appendix J: Rubric for Scoring Summative Standardized Assessment of Evidence-Based Practice.....	222
Appendix K: Initial and Final Versions of the Familiarity with the Process of Evidence-Based Practice in Education Scale with Summary Form from Expert Review Committee.....	223
Appendix L: Quiz on Basic Information Literacy Skills (Iteration 5).....	227
Appendix M: Rubric for Scoring Education Research Project (Iteration 5).....	232
Appendix N: Sample Table for Reporting Results from Generalizability Theory Analysis .....	234
Appendix O: Sample Data Setup for Generalizability Theory Analysis .....	235
Appendix P: Standardized Performance Assessment Follow-up Interview Protocol.....	236
Appendix Q: Copyright Permissions .....	236



## List of Tables

Table 1:	Constitutive Definitions of Design-Based Research.....	74
Table 2:	Contrasts Between Design Experiments and Laboratory Studies.....	76
Table 3:	Theoretical Framework for the Process of Evidence Based Practice .....	84
Table 4:	Conceptual Framework for the Process of Evidence Based Practice .....	90
Table 5:	Correspondence Between Tasks in the ERP and the Five-Step EBP Framework .....	95
Table 6:	Results Reported by Section from Information Literacy Questionnaire Item 8D.....	98
Table 7:	Revised Rubric for the First Four Tasks of the Education Research Project...99	
Table 8:	Descriptive Statistics for Sections A and B on the Three Indices for the Education Research Project .....	100
Table 9:	Revised Education Research Project Used in Iteration Four .....	110
Table 10:	Education Research Project- Residency Version Used in Iteration Five.....	113
Table 11:	Test Retest Reliability Coefficients for the Information Literacy Questionnaire Items .....	128
Table 12:	Changes in Pretest and Posttest Scores on Information Literacy Questionnaire Items 1, 2, 4, and 11 for Iterations 1-5 .....	130
Table 13:	Changes in Pretest and Posttest Scores on Items 3, 5, 7, 8, 9, and 10 for Iterations 1-5 .....	131
Table 14:	Changes in Pretest and Posttest Scores on Items 6A, 6B, 6C, and 6F for Iterations 1-5 .....	132
Table 15:	Changes in Pretest and Posttest Scores on Items 6D, and 6E for Iterations 1-5 .....	133

Table 16: Item Level Descriptive Statistics for the Familiarity with the Process of Evidence-Based Practice in Education .....	137
Table 17: Research Topics and Number of Residents Researching the Topic for the Education Research Project- Residency Version.....	139
Table 18: Descriptive Statistics for Scoring of the Education Research Project- Residency Version .....	140
Table 19: Relationships of Original Codes to Dichotomous Success Categories for Four Themes from the Research Application Essays of Iterations One and Five .....	141
Table 20: Descriptive Statistics for Item-Level Scores from the Standardized Performance Assessment .....	143
Table 21: Sources of Variance in the Three-Facet Generalizability Study.....	144
Table 22: Relative and Absolute Generalizability Coefficients from Decision Studies in which Testing Conditions of Raters, Tasks, and Occasions were Manipulated .....	145

## **List of Figures**

Figure 1:	The continuum of the integration of research and practice from the perspective of the practitioner (teacher). .....	10
Figure 2:	Three levels of analysis that can be used to address the problem of lack of equity and lack of achievement in mathematics .....	11
Figure 3:	Theoretical Framework for the Process of Evidence Based Practice .....	23
Figure 4:	Overview of the iterations of the DBR agenda.....	93
Figure 5:	Overview of the five iterations of data analysis for the application to research essays (Iteration 1).....	102
Figure 6:	Concept table used to prepare for the Education Research Project .....	106
Figure 7:	Graphical representation of reliable change index scores for eleven of the participants in Iteration 5 .....	138
Figure 8:	Box plot of reliable change index scores for eleven of the participants in Iteration 5 .....	138
Figure 9:	A comparison of Iteration 1 and Iteration 5 participant success rates on communication of four categories related to process of applying research to practice in the research application essays from the Education Research Project .....	142
Figure 10:	The relationship between the number of words and the score for Iteration Five standardized performance assessments .....	144

## **Abstract**

Persistent lack of mathematics achievement and disparity in achievement has led to the publication of research findings related to equitable teaching practices. Although the publication of such research provides insights about approaches for potentially increasing equity in mathematics education, teachers must be able to apply what has been learned from these studies to their classroom teaching practices. Despite the widespread expectation that teachers use research-supported teaching strategies to meet the needs of their diverse classrooms, the research to practice gap persists. Little research is currently available to guide mathematics teacher educators in how to prepare future teachers to apply research to teaching practices.

Inspired by advancements in social work and other health-related fields, this study departed from the standard approach of preparing teachers to utilize specific, research-based teaching strategies to preparing teachers to engage in the *meta*-process of applying research to practice. This meta-process has been defined by the health-related disciplines as the process of evidence-based practice (EBP). This process is explicated in a conceptual framework that is composed of the following five steps. The practitioner (1) formulates an answerable practice question, (2) searches for the best research evidence, (3) critically appraises the evidence, (4) selects the best intervention for a specific practice context, and (5) evaluates the outcome of the intervention.

The purpose of this study was to examine the process of preparing preservice elementary teachers of mathematics to engage in the five-step process of EBP. Because

this process, which can be conceptualized as a routine of practice, has not been identified for the field of mathematics education previously, it was examined using a design-based research (DBR) methodological approach. There were two objectives to the study: (1) to create an empirically tested teaching intervention that mathematics teacher educators can use to prepare preservice teachers to apply research to teaching practice and (2) to create a system of assessment that supports the teaching of this intervention.

The study involved five iterations of the DBR process that permitted the intervention to be evaluated and revised after each iteration. Although each iteration is discussed, this study focuses primarily on the process used in the fifth iteration of the DBR process. This iteration took place in the context of a mathematics methods course in a clinically-rich, undergraduate residency program for initial preparation of elementary school teachers. The twelve participants were simultaneously enrolled in the methods course and embedded in co-teaching assignments at an elementary school.

The intervention to prepare teachers to engage in EBP included two workshops that were co-facilitated by an education librarian and a mathematics teacher educator and a semester-long Education Research Project. The project required participants to identify a problem of practice related to teaching or learning mathematics, find relevant research to address that problem, create an intervention to apply the research findings to classroom instruction, implement that intervention, and collect data to evaluate the effectiveness of the designed intervention.

Instruments used to collect data included: (1) a self-report Information Literacy Questionnaire, (2) a self-report Familiarity with the Process of Evidence-Based Practice in Education Scale, (3) the Education Research Project report, and (4) a standardized

performance assessment. The standardized performance assessment was used to assess beginning proficiency with the process of EBP. Generalizability theory was used to evaluate the reliability of the system created for the standardized performance assessment. The system that included three raters, two tasks, and two scoring occasions was found to be fairly reliable (absolute generalizability coefficient = .81).

Results from this study revealed that participants were more successful at creating implementation plans and linking those plans to research than they were at modifying their plans to meet the needs of specific students or evaluating their research implementation. This study contributes to both research and mathematics education communities' understandings about the potential of EBP as a high-leverage routine of practice and the use of generalizability theory in the creation of a reliable assessment to evaluate this routine of practice. This study documents the complexity of the process of linking research to practice and provides an empirically tested conceptual framework for preparing preservice teachers to engage in this complex practice.

## Chapter 1: Introduction

“There are large, persistent disparities in mathematics achievement related to race and income—disparities that are not only devastating for individuals and families but also project poorly for the nation’s future, given the youthfulness and high growth rates of the largest minority populations.” (National Mathematics Advisory Panel, 2008, p.xii)

### **Problem Statement: Persistent Disparities in Mathematics Achievement**

In the field of engineering, the bottom line is the functionality of the design. In medicine, the bottom line is the health of the patient, and in education, the bottom line is the achievement of the student. When student achievement is seen as lacking, teachers, parents, politicians, and most of society begin looking for ways to improve the bottom line. Today, in the United States, there is a widely held perception that student achievement in mathematics is lacking.

Such a perception often develops as a result of poor performance on exams or poor outcomes related to job eligibility or performance. One does not need to look far to find evidence of poor performance on mathematics assessments. To illustrate this, I will present a few specific examples, but many more could be cited. At the state level, in 2012, 40% of all Floridian 4<sup>th</sup> grade students who took the Florida Comprehensive Achievement Test scored below the proficient level (Florida Department of Education, 2012). At the national level, in 2009, about 61% of US 4<sup>th</sup> graders scored below the proficient level on the National Assessment of Educational Progress (NAEP) (Aud, Fox, & Kewal Ramani, 2010). At the international level on the 2007 Trends in International Mathematics and Science Studies (TIMSS), 10% of US 4<sup>th</sup> graders scored at the

advanced level, compared with 41% of 4<sup>th</sup> graders from Singapore (Gonzales et al., 2008).

When looking at any one indicator of mathematics achievement, there are many factors to consider in the interpretation of the score. One must reflect on the construct that was being measured and on the instrument that was used in the measurement. Students might score low on an exam that measured their ability to use a slide rule yet such an exam would be almost meaningless in the digital era of the twenty-first century. Although caution should always be exercised in the interpretation of achievement scores, a multitude of low scores provides a more robust case for the existence of a problem in student achievement in mathematics in the U. S.

The case that I am building for concern over low mathematics achievement in the US is strengthened exponentially if one looks at the disparity in achievement of subgroups of US students. Again, for the sake of this argument I present findings of disparity for racial subgroups, but similar findings exist when one examines economic and sociolinguistic subgroups. Looking again at the 2009 NAEP scores, about 61% of US 4<sup>th</sup> graders scored below the proficient level on the National Assessment of Educational Progress (NAEP) (Aud, Fox, & Kewal Ramani, 2010). Although the lack of proficiency alone ought to be concerning to mathematics educators, the disparity in proficiency when scores are disaggregated by race ought to be alarming. While 39% of Asian/Pacific Islanders scored below proficient, 84% of Black students scored below proficient. This disparity is only exacerbated with additional schooling. At 12<sup>th</sup> grade, about 95% of African American students scored below the proficient level as compared



with 64% of Asian/Pacific Islander students. That only 5% of Black 12<sup>th</sup> graders scored at or above the proficient level is cause for alarm!

I have discussed an educational problem in mathematics achievement at the student level: there is a problem with overall lack of achievement, and there is disparity in achievement among racial subgroups. Through one lens, this is a student-level problem, but this problem can also be addressed at the teacher level. The decision to shift the level of analysis from student level to teacher level is ubiquitous in public and professional discourse about education. This shift in focus has been driven, in part, by statistical models of student achievement that have indicated that teachers can have large effects on student learning (Nye, Konstantopoulos, & Hedges, 2004) and that these large effects can have lasting impact on achievement (Gordon, Kane & Staiger, 2006; Rowan, Correnti, & Miller, 2002; Sanders & Rivers, 1996). Models such as the one produced by Rivkin, Hanushek, and Kain (2005) indicated that teachers' effectiveness had a greater impact on student learning than any other factor that can be controlled by the school system. This provides a strong rationale to focus attention and resources on improvements at the teacher level to realize improvements at the level of student achievement.

Similar to the caution one must exercise when interpreting any one measure of student achievement, one must also use caution when interpreting the statistical models that identify teacher effects upon student achievement. However, just as there is a pattern of lack of and disparity in student achievement across many different assessments, there is also a pattern of significant and meaningful effects of effective teachers on student

learning. Therefore, the current emphasis on teacher effectiveness is an appropriate level at which to address the problem of student achievement in mathematics.

When the level of analysis changes from the student to teacher, so do the ways in which one describes the achievement problem. When lack of student achievement is examined through a teacher-level lens, one begins to ask why large numbers of teachers are not facilitating equitable and successful mathematics learning environments. In the end, one must ask what behaviors teachers are or are not engaging in that create an environment where students, particularly minority students, are falling through the cracks and failing to master essential mathematics concepts. However, prior to addressing this question, it is prudent to understand a little more about the diversity that exists in today's public school classrooms.

Today's US public school teachers are leading incredibly diverse classrooms. In a US Department of Education report (Aud et al., 2010), the following facts were presented with respect to the growing diversity found in US public school classrooms. From 2000 to 2007, the proportion of White public school students dropped from 61% to 56% and enrollment of Hispanic students increased from 17% to 21%. In addition to being racially diverse, US public schools are linguistically diverse. In 2007, 21% of elementary and secondary students spoke a language other than English at home. Furthermore, public schools also have significant diversity related to learning exceptionalities. Again, in 2007, 9% of 6- to 21-year-olds were served under the Individuals with Disabilities Education Act (IDEA). In summary, the average public school teacher walks into a classroom where close to half of the children are racial

minorities, one-fifth speak a language other than English at home, and one-tenth have a learning exceptionality that is covered under IDEA.

Society asks, and indeed laws mandate, that this one teacher teaches with equity to each student in this very diverse classroom. Based upon the achievement results, it is clear that teachers have been unable, given their current resources, to meet this challenge. After all, how does one teacher know how to facilitate an equitable learning environment with a classroom whose members include students with psychological and emotional disorders, learning exceptionalities, several different home languages and levels of English language proficiency, unique histories of learning mathematics, and unique cultural identities? How does a teacher know how to teach with equity to a student whose needs the teacher has not encountered before or with whom the teacher has not found success before?

Education research is a potential source of knowledge that teachers could draw upon in order to begin to meet the needs of such a diverse body of students. Perhaps, through education research, teachers could learn about specific practices that have the potential to meet the needs of the diverse student population. The potential exists and has been recognized by educational organizations around the world, but, as the next section of the introduction discusses, there currently exists a wide gap between education research and teaching practices.

### **The Challenge of Linking Research to Teaching Practice**

Over the past decade, there has been international interest in how to link the results of education research to the decisions that impact teaching and learning in the classroom. This interest has been reflected in the national education agendas of countries

such as Canada, Denmark, New Zealand, Singapore, and the United Kingdom (OECD, 2007). A working document by the Commission of the European Communities (2007) stated that evidence-based practices and policies should drive reform in education, and the document provided an outline for reform based on the creation, application, and mediation of knowledge in the classroom. Despite sustained, international interest in applying research to practice, there is also widespread agreement that a gap persists between research and practice. In other words, teaching practices have not been consistently improved by research on the effectiveness of those practices (OECD, 2009).

The research to practice gap has also been the subject of much attention in the United States. The federal legislation known as No Child Left Behind (2002) referred to “scientifically-based practices” more than one hundred times. This legal mandate for the connection between research and practice was supported by the creation of the What Works Clearinghouse and the resources promoted by the Institute for Education Sciences (IES). One decade after No Child Left Behind, the federal initiative called Race to the Top tied the awarding of federal education funds to the use of evidence-based practices. Furthermore, the *Blueprint for Reform: The Reauthorization of the Elementary and Secondary Education Act* (U.S. Department of Education, 2010) clearly tied distribution of federal funds to the use of evidence-based practices. The following language from that document illustrates the strongest commitment toward encouraging and mandating the use of evidence-based practices in education.

We will support states, districts, school leaders, and teachers in implementing a more complete education through improved professional development and evidence-based instructional models and supports. (p. 4)

States will be required to develop comprehensive, evidence-based, preK–12 literacy plans and to align federal, state, and local funds to provide high-quality literacy instruction. (p. 26)

This program builds on the i3 program launched through the American Recovery and Reinvestment Act of 2009, and will provide additional competitive grants to expand the implementation of, and investment in, innovative and evidence-based practices, programs, and strategies that significantly improve student outcomes. The Secretary will use a rigorous, three-tiered evidence framework that directs the highest levels of funding to programs with the strongest evidence, and also provides significant support for promising programs that are willing to undergo rigorous evaluation. (p. 36)

These passages are a few among many that illustrate how the language of evidence-based practice has permeated legislative discourse about education. To this day, national conversations continue about how to achieve the illusive connection between research and practice at the level of teaching and learning in the classroom.

Just as there has been interest in linking research to practice in the broad field of education in the United States, there has been interest in this link within specific content areas as well. The field of mathematics education is no exception. Over the past 10 years, the National Council of Teachers of Mathematics (NCTM) has engaged in sustained dialogue about this link. In 2004, NCTM established a Linking Research to Practice Task Force (LRPTF). This task force's recommendations continue to guide NCTM's actions toward closing the research to practice gap. NCTM's Research Committee (Gutstein et

al., 2005) explicitly referred to the connection between research and practice as a key element to achieving equity in mathematics education. They reported:

Research impacting mathematics education is increasingly important in the decision-making that characterizes the day-to-day work of school district personnel, classroom teachers, and policymakers. In response to these needs, NCTM has adopted as a major goal the linking of research and practice, for example, in convening the Research Catalyst Conference. This commitment couples the goal of helping practitioners understand and use research with that of helping researchers understand and study practitioners' most critical questions. An equity focus for research is responsive to practitioners' needs, reflective of NCTM's longstanding commitment to equity, ideal as a site for linking research and practice, *and* the right thing to do. (p. 99)

The above statement reflects the urgency of the need to link research to practice as a condition that paves the way for the achievement of equity in mathematics education. Teachers cannot rely solely on their own experiences or the experiences of their colleagues to guide their decision-making for the creation of an equitable classroom. Perhaps only with the knowledge gained from mathematics education research can teachers hope to gather the information needed to be equipped to create equitable learning environments. In light of NCTM's commitment to equity in mathematics education, one begins to see why this organization has also made a strong commitment toward bridging the gap between research and practice.

The report, *Linking Research to Practice: The NCTM Research Agenda Conference Report* (Arbaugh et al., 2010), is an example of NCTM's commitment to

bridge the research to practice gap. This report was the culmination of a working conference that established a mathematics education research agenda informed by the experiences and questions of K-12 mathematics teachers. The report presented 25 key research questions that reflected the needs of teachers. In addition, *Linking Research to Practice* also set forth a vision of what the full integration of research and practice would look like from the perspective of a researcher and from the perspective of a teacher.

Figure 1 is a reproduction of the figure that illustrates the integration of research and practice from the perspective of the practicing teacher. At the bottom of the figure, where integration is lowest, the teacher does not interact at all with education research. At the top of the figure, where integration is highest, the teacher fully engages in the research cycle. The authors of *Linking Research to Practice* noted that the fourth level in this illustration represents a shift in the connection between research and practice (Arbaugh et al., 2010). It is at this level that a teacher applies the results of research to teaching practice.

Although NCTM has publicly stated the goal for teachers to read and apply research regularly to their teaching practice, there is evidence to suggest that teachers are not yet doing this (OECD, 2009). In fact, teachers have reported lack of interest in education research. Some teachers have claimed that research does not address their concerns about teaching (Gore & Gitlin, 2004) and that research results can be used to show or support anything (Boardman, Arguelles, Vaugh, Hughes, & Klinger, 2005). It is the disconnect between the classroom teacher and education research that is the central problem that this study will address.

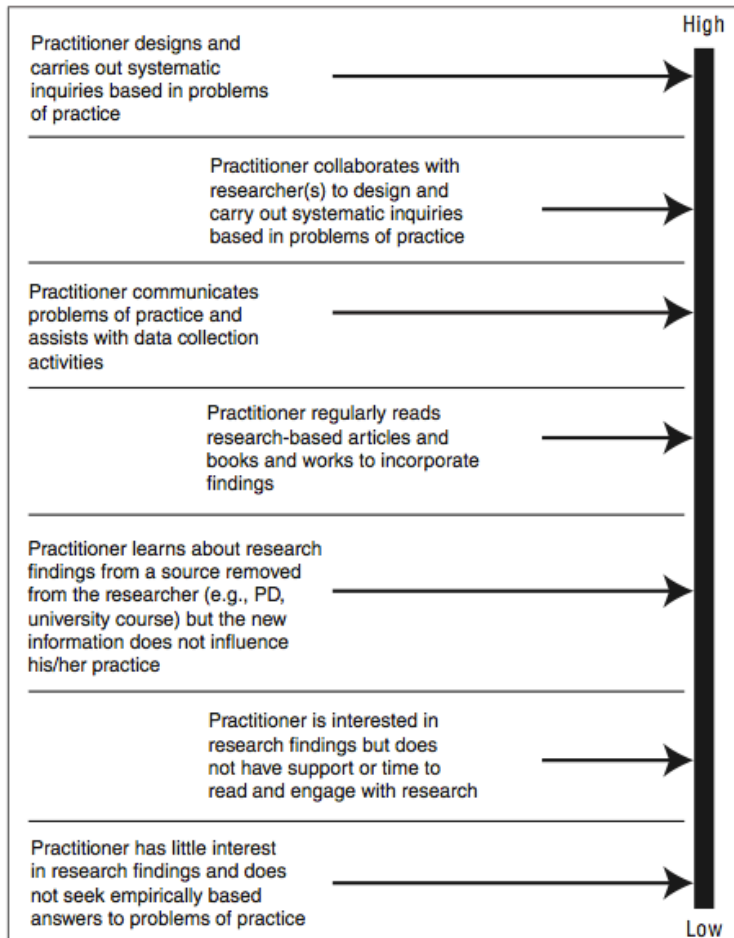
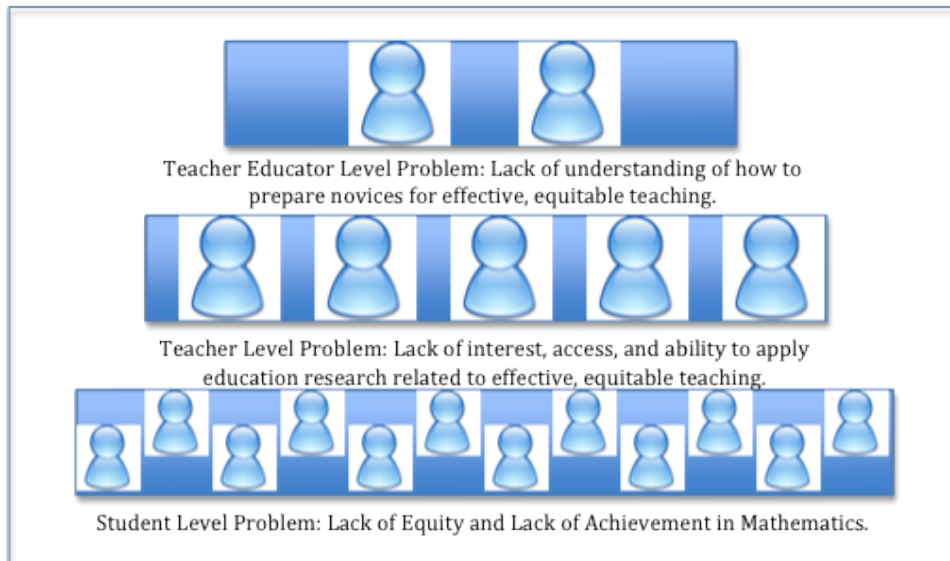


Figure 1. The continuum of the integration of research and practice from the perspective of the practitioner (teacher). Adapted from *Linking Research to Practice: The NCTM Research Agenda Conference Report* by F. Arbaugh, B. Herbel-Eisenmann, N. Ramirez, E. Knuth, H. Kranendonk, and J. R. Quander, 2010, p. 33, Reston, VA, National Council of Teachers of Mathematics. Copyright 2010 by the National Council of Teachers of Mathematics. Adapted with permission.

If we continue to address lack of student achievement and the gap between research and practice at the teacher level, we will be able to evaluate the practices of individual teachers but will be unable to address levers of widespread change. However, if we address the level of teacher education, then we will be able to ask questions about how teachers are currently educated and how teacher education programs might be changed so that teachers are better prepared to teach in ways that are effective for *all* students. Figure 2 provides a schematic diagram that summarizes the three potential



levels of analysis and the phrasing of the main problem at each level. The next section discusses some previous approaches to closing the research to practice gap.



*Figure 2.* Three levels of analysis that can be used to address the problem of lack of equity and lack of achievement in mathematics.

### **Previous Approaches to Linking Research to Practice**

Education researchers, administrators, and governmental ministries from all over the world have tried to address the gap between research and practice. Teacher educators have attempted to prepare teachers to use specific instructional practices that research has shown to be effective. Two examples of failed attempts to encourage mathematics teachers to apply research to practice are presented below as a means to illustrate the most typical approaches to closing the gap between research and practice.

By the mid-1990's there was robust evidence that use of cooperative learning (CL) in mathematics classrooms had significant positive impact upon student achievement. Professional development programs on CL proliferated. Teachers, both preservice and inservice, were taught to use this strategy that had been shown to be

effective. Because a large number of teachers were taught to use this research-based strategy, one would assume that the strategy would find its way into the classroom. Antil, Jenkins, Wayne, and Vadasy (1998) found that of 93% of teachers ( $n=85$ ) did indeed claim to use CL in their classrooms. However, observations of these teachers' classrooms revealed that only 5% (up to 24% with a more flexible understanding of CL) of teachers actually used CL in a way that was consistent with the research.

The discrepancy between teacher perception of the use of research-based teaching practices and observable teaching behavior was also found in the TIMSS Video Studies (Stiegler & Hiebert, 2009). An examination of videos of mathematics lessons taught in 1995 revealed that US teachers facilitated learning environments where students worked on mathematical procedures and skills but were denied the opportunity to engage in the rich work of sense-making and non-routine problem solving. Stiegler and Hiebert (2009) reported that, although the US engaged in "massive efforts" to improve and reform math education during the years between 1995 and 1999 and, although teachers professed that their teaching had changed in substantial ways due to the research that was presented to them, in fact no changes in instructional practices were observed. In the 1999 videos, students were still spending most of their time reviewing material that had been previously taught, and teachers enacted lessons that were focused on procedural fluency rather than deeper conceptual understanding.

### **The Process of Evidence-Based Practice**

The two studies discussed in the previous section illustrate some of the difficulties associated with transferring successfully the knowledge of specific evidence-based practices to the real-time teaching actions of a mathematics classroom teacher. This

approach requires teachers to “learn” lists of practices that are supported by evidence. Such lists of practices are easily forgotten in the press of real-time teaching (Cordingley, 2008). Instead of teaching teachers how to use specific evidence-based practices, it may be more efficient to teach teachers how to engage in the *process* of applying research to practice. This would change the focus away from specific teaching actions and towards a broad pattern of actions. This shift in focus would be similar to the shift made by cognitive psychologists away from teaching students lists of learning strategies and toward teaching *metacognitive* skills. The research to practice gap might be reduced if teachers were taught the *meta* skill, or routine of practice, of applying research to classroom teaching.

I began this chapter with a discussion about the lack of mathematics achievement and disparity in mathematics achievement for K-12 public school students in the US. Because effective teaching has been shown to have the greatest influence on student achievement, among variables that school districts can manipulate, public attention has focused on improving teaching as a way to improve mathematics achievement. How does one improve teaching? This is, perhaps, the most salient and elusive question in education today. An often-suggested strategy is to have teachers use teaching practices that have been supported by evidence from well-constructed research studies. Given the diversity in the public school classroom, it seems reasonable that teachers need information from education research in order to facilitate equitable learning environments. This is likely why education legislation has mandated that teachers use practices that are supported by research. However, despite such mandates for teachers to apply research to their teaching practices, and despite efforts to teach teachers about

specific evidence-based practices, the gap persists between research and teaching practice.

The gap between research and practice might better be addressed through a change of focus in teacher education. Instead of trying to teach teachers about a myriad of specific research-based strategies, teacher educators would prepare teachers to engage in a *routine* or *meta*-strategy that would problematize the research to practice gap and provide teachers with a roadmap for making connections between research and practice.

### **Conceptual Framework**

The decision to address the research to practice gap by training teachers (practitioners) to engage in a meta-routine of applying research to practice has been informed by similar work in the fields of medicine, psychology, and social work, among others. In the field of medicine, the term *evidence-based practice* (EBP) has been defined as “the integration of best research evidence with clinical expertise and patient values” (Sackett, Straus, Richardson, Rosenberg, & Haynes, 2000, p. 1). In the context of social work, Rubin (2008) further explicated the definition of EBP into these two overarching perspectives: 1) EBP is a *process* that involves finding and evaluating research in the context of making practice decisions and 2) EBP is a way to designate certain practices as empirically supported. In the field of social work, this distinction has clarified for researchers and practitioners the difference between designating a treatment as an EBP and designating the *process* of practice in which a clinician engages as EBP. By clearly defining the terminology, Rubin (2008) paved the way for more constructive dialogue on EBP and interventions that strengthened social workers’ abilities to engage in the process of EBP (Rubin & Parish, 2007, 2011).

In addition to defining EBP as a *process*, social work researchers (Mullen, 2004, 2006; Rubin, 2008; Shlonsky & Gibbs, 2004; Thyer, 2006) have also explicated a conceptual framework that outlines the five steps that are necessary to engage in EBP:

1. Formulate an answerable practice question.
2. Search for the best research evidence.
3. Critically appraise the research evidence.
4. Select the best intervention after integrating the research evidence with client characteristics, preferences, and values.
5. Evaluate the outcome of this practice decision.

This five-step framework has provided guidance for training social workers on how to engage in the process of EBP (e.g., Parish & Rubin, 2011; Rubin, 2008). Each of these steps has been defined, operationalized, and used to train practitioners in the process of implementing research in practice.

Although the process of EBP has been operationalized and researched in many fields (e.g., medicine, psychology, occupational therapy), the field of social work has been chosen as a model for research on EBP training for preservice teachers for two reasons: (a) social work researchers have looked specifically at the preservice training of social work practitioners and (b) the educational preparation of social workers is conducted mainly at the level of the undergraduate and master's degrees (in contrast to medicine which requires doctoral education). Because of similarity in preparation for social workers and teachers and because the five-step framework has been useful in the training of social workers, this framework has been selected to guide the training of preservice teachers. A learning module will be developed based on this framework to

prepare preservice teachers to engage in the routine of applying research to classroom teaching. The goal is to develop a meta-routine that would prepare teachers to apply research to their teaching practice.

### **Research Purpose and Study Design**

The overarching purpose of this study is to determine how to prepare preservice elementary teachers of mathematics to engage in the routine of applying research to practice. Because this routine of practice has not been defined for the field of mathematics education, it will be explored using a design-based research (DBR) investigation. Confrey (2006) defined DBR in this way:

A design study is an extended investigation of educational interactions provoked by use of a carefully sequenced and typically novel set of designed curricular tasks studying how some conceptual field, or set of proficiencies and interests, are learned through interactions among learners and with the guidance of an instructor or form of tutor. The study seeks to document what resources and prior knowledge the students bring to the task, how students and teachers interact, how records and inscriptions are created, how conceptions emerge and change, what resources are used, and how teaching is accomplished over the course of instruction, by studying student work, video records, and classroom assessments.

(p.2)

The current study fits well with Confrey's definition of DBR. The five-step conceptual framework for EBP was used to create a learning module, a set of "carefully sequenced and novel" instructional activities, aimed at preparing preservice teachers to apply research to practice. The implementation of this learning module was examined over a

series of iterations. In each iteration the design of the learning module, the conditions of the implementation, and the evidence of learning by the preservice teachers were evaluated. Examination of the evidence of learning required the development of assessments that measure this routine of practice. The analysis of each iteration was then used to refine the learning module and the accompanying assessments. Revised learning modules were then implemented and the cycle of implementation, analysis, and revision continued.

To achieve the goal of preparing preservice teachers to engage in EBP, there are two distinguishable objectives for this DBR study: 1) to use the conceptual, five-step framework for the process of EBP to create a product, a teaching intervention, that mathematics teacher educators can use to prepare preservice teachers to apply education research to teaching practice; and 2) to create a system of assessment that supports the teaching intervention. Ultimately, this study attempts to answer the question of how teachers can learn the meta-practice of applying research to their teaching. As a part of this research agenda, the five-step conceptual framework of the process of evidence-based practice must be explicated and shown to be applicable to the practice of teaching. If found to be applicable, then the developed intervention can be used by teacher educators to prepare preservice teachers to engage in the meta-practice of applying research to practice. Further, the developed system of assessment can be used along with the intervention to evaluate its effectiveness.

## **Research Questions**

Each iteration of a DBR study has research questions that are unique to that iteration. Because this study was based on the fifth iteration of the overall DBR agenda, I provide the research questions that were addressed in this iteration:

- 1) After experiencing the intervention, to what extent did the participating preservice teachers demonstrate the ability to apply research to their classroom teaching?
- 2) To what extent did preservice teachers' intentions to apply research to future teaching change from the beginning to the end of the intervention?
- 3) What were the differences in a) ability to apply research to practice and b) intention to apply research to practice between participants in the fifth iteration and participants in previous iterations?
- 4) How reliable were scores from the system of assessment that was used to assess preservice teachers' abilities to apply research to teaching?

## **Delimitations and Study Significance**

Fundamentally, this study addresses teacher training. There are two sites of teacher training: inservice training (professional development) and preservice training (teacher preparation programs). This study is delimited to address only preservice training that occurs as a part of an undergraduate elementary education teacher preparation program. Studies that explore inservice teacher training with the five-step model for evidence-based practice are also needed but are beyond the scope of this study.

This study is housed within the larger context of teams of researchers who are working to define common routines of practice for mathematics teaching that can be taught as part of a teacher preparation program (Ball, Sleep, Boerst, & Bass, 2009;



Boerst, Sleep, Ball, & Bass, 2011; Franke, Grossman, Hatch, Richert, & Schultz, 2006; Kazemi, Lampert, & Ghouseini, 2007). Because of this, the vocabulary and guidelines developed by this group (Chapter 2 will provide a review of work related to routines of practice) will be used. Using common language and common frameworks will enable the field of mathematics education to build cumulative knowledge and make forward progress. This study will contribute a new high-leverage routine of practice, that of the process of evidence-based practice, to this developing body of literature.

Because the process of evidence-based practice has not yet been defined as a routine of practice for the field of mathematics teacher preparation, there is much exploratory research that must be conducted in order to define the process and establish a system of assessment for the practice. This design-based research study is exploratory in nature. As such, this study has not been designed to make claims about generalizability to a larger population other than the samples that comprise this study. Furthermore, this study is not designed to isolate and describe the effects of individual variables related to teaching preservice teachers how to engage in the process of evidence-based practice. Rather, in accordance with the goals of design-based research, this study is designed to evaluate, simultaneously, an intervention and the conditions that give rise to that intervention.

This study is significant because it will produce a product—a learning module—that can be used to prepare preservice teachers to link research to practice in education in general, and in mathematics education in particular. In this way, the study responds to the call to define and develop high-leverage routines of practice that can be addressed in teacher preparation. This study will also produce a system of assessment that can be used

to assess the high-leverage routine of evidence-based practice. This system will address the need to develop systems of assessment for performance tasks (Hill, Charalambous, & Kraft, 2012).

### **Limitations**

In general this design-based research study had inherent limitations because large amounts of data were produced during the study, and there was the potential for the results to be influenced by researcher bias in the selection and interpretation of the data. In addition, small sample size was a limitation for each of the iterations. Even though this study did not intend to isolate variables or speak to causal relationships among variables, history and extraneous variables remained as threats to validity. As Chapter 3 details, every effort was made to mitigate these threats insofar as the design allowed.

## **Chapter 2: Literature Review**

The purpose of this chapter is to provide a review of the literature that guided the design of the study to be introduced in Chapter Three. I began the review process with the goal of understanding how preservice teachers of mathematics learn to engage in evidence-based practice (EBP). Because researchers in the field of social work have defined a conceptual framework for the process of EBP and studied its use in the preparation of social work practitioners, I first reviewed the EBP literature in this field. Then I reviewed teacher preparation literature to understand what is known about linking research to practice within the specific environment of teacher preparation.

Having gathered knowledge about EBP and the process of linking research to practice in teacher preparation, I reviewed literature on routines of practice in mathematics teacher preparation. This review was conducted to understand how EBP in teacher preparation relates to the broader movement of identifying high-leverage teaching practices.

A common theme from the first three areas of review was the importance of the use of performance assessments to measure the learning of new processes. For this reason, I also reviewed literature on performance assessments. Finally, design-based research (DBR) emerged as a potential approach to research design for studying novel educational interventions. I reviewed DBR literature in order to learn more about the indications of quality for this methodological approach. After presenting the findings of

these five areas of literature, I conclude this chapter with discussion of the implications of the literature review for the design of the study introduced in the next chapter.

### **Preparing Practitioners for Evidence-Based Practice**

In this section I discuss what is known about preparing social work practitioners to engage in the process of evidence-based practice (EBP). Because research on EBP in medicine and other health-related fields has been foundational to EBP research in the field of social work, important studies in these health-related fields will be included in this review as well.

**Defining evidence-based practice.** Evidence-based practice has been defined for the medical field as the “conscientious, explicit, judicious use of current best evidence in making decisions about the care of individual patients” (Sackett, Rosenberg, Gray, Haynes, & Richardson, 1996, p. 71) and “the integration of best research evidence with clinical expertise and patient values” (Sackett, Straus, Richardson, Rosenberg, & Haynes, 2000, p. 1). Scatterfield, Spring, Brownson, Mullen, Newhouse, & Whitlock (2009) created a transdisciplinary model of EBP (Figure 3) that was based upon research from the following fields: medicine, nursing, psychology, social work, and public health. Figure 3 shows how the transdisciplinary model takes the elements of the definition by Sackett et al. and places those elements within the environmental and organizational context.

Within the field of social work, the term “evidence-based practice” has been used to refer both to specific interventions supported by research and to the *process* in which practitioners use the best evidence available to make decisions about the care of specific

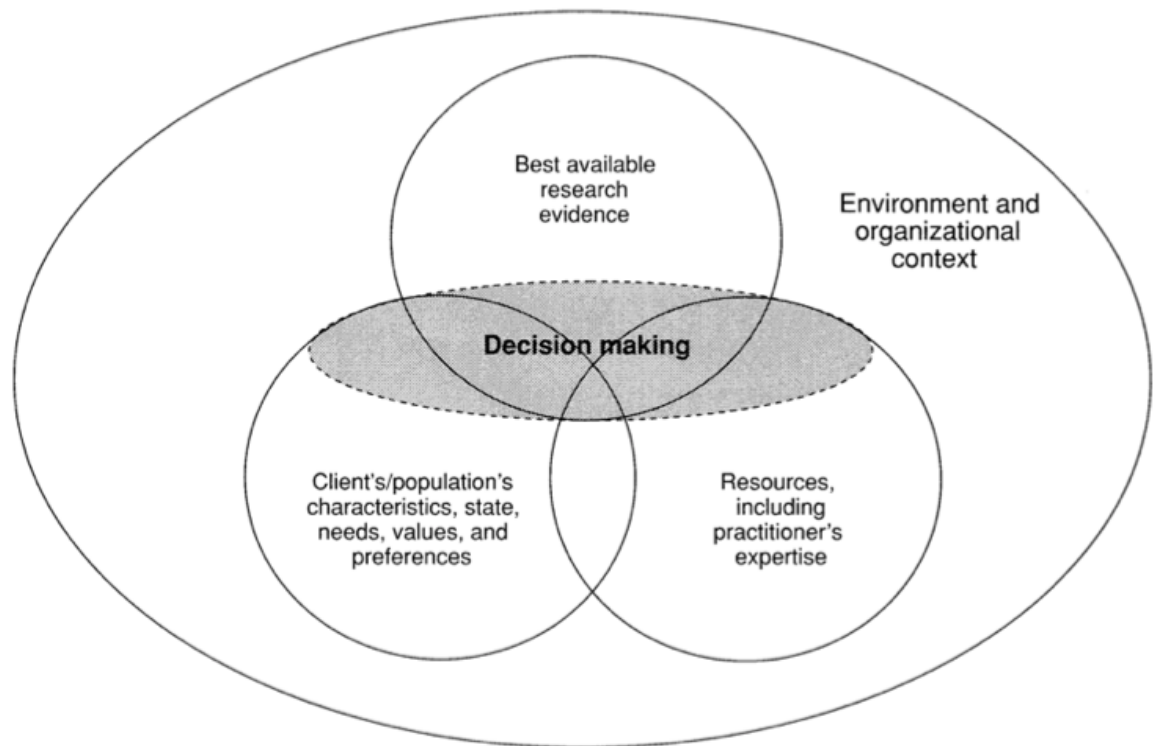


Figure 3. Interdisciplinary model of EBP. Reprinted from “Toward a Transdisciplinary Model of Evidence-Based Practice” by J. M. Scatterfield et al., 2009, *Milbank Quarterly*, 87, pp. 368-390. Copyright 2009 by Wiley. Reprinted with permission.

clients (Rubin, 2007, 2008; Rubin & Parrish, 2007).

Social work researchers have developed a five-step framework, first introduced in the field of medicine (Sackett & Rosenberg, 1995), to operationalize the process of EBP for practitioners. These steps are: (1) Formulate an answerable practice question, (2) Search for the best research evidence, (3) Critically appraise the research evidence, (4) Select the best intervention after integrating the research evidence with client characteristics, preferences, and values, and (5) Evaluate the outcome of this practice decision (Gibbs & Gambrill, 2002; Parish & Rubin, 2011; Rubin, 2008; Shlonsky & Gibbs, 2004; Thyer, 2006). This framework is compatible with the understanding that

EBP is an active *process* in which a practitioner engages. Although this process view of EBP is central to most of the literature discussed in this section, it is not the only interpretation of the term EBP. Rubin and Parrish (2007) documented that about one quarter of social work faculty ( $N = 973$ ) who responded to a national, online survey viewed EBP as a term used to indicate a particular treatment or intervention as empirically supported.

**Outcomes of training on EBP.** It is possible to evaluate training on the process of EBP by assessing clinical outcomes and/or educational outcomes (Khan & Coomarasamy, 2006). Clinical outcomes are defined by the effect that the EBP training has upon the patient or client. Educational outcomes are defined by the learning achieved by the practitioner (or student) who underwent the EBP training.

*Clinical outcomes.* From the inception of training on EBP in medicine, researchers have acknowledged the difficulty of assessing clinical outcomes. The Evidence-Based Medicine Working Group (1992, p. 2424) stated that “proof” of the efficacy of training in evidence-based medicine would not be achievable given the impracticalities of the long-term randomized trials that would be needed to compare traditional medical practice to evidence-based medical practice. Hatala and Guyatt (2002) further explicated this difficulty by pointing to the challenge of defining and measuring the change in patient outcomes that would be expected as a result of EBP training for practitioners. However, Hatala and Guyatt also pointed out that this difficulty is not unique to training on EBP. Just as there has not been convincing proof that EBP improves patient outcomes, neither has there been proof that the traditional

approach to medical teaching on topics, such as conducting a physical exam, actually improves patient outcomes.

Furthermore, there is some evidence that, in the absence of training on EBP, practitioners fail to use the best available treatments or interventions. In a now dated study, Schuster, McGlynn, and Brook (1998) found widespread evidence of inappropriate and even harmful care for patients in the United States medical system. Choudry, Fletcher, and Soumerai (2005) conducted a systematic review ( $N = 62$  articles) of the relationship between the number of years a physician has been in practice and the quality of patient care. They found a negative relationship between these constructs and hypothesized that the older physicians were not keeping up with the most effective medical practices. Pignotti and Thyer (2012) surveyed 400 clinical social workers to determine their use of empirically supported treatments (ESTs) versus use of novel unsupported therapies (NUSTs). They found that 75% of the sample used at least one NUST. In addition, the social workers rated clinical experience more highly than research evidence as a reason for selecting a particular intervention. Although the three studies cited do not provide evidence of the effectiveness of EBP training, they demonstrate the dangers of the lack of practitioner engagement in the process of EBP.

Despite the difficulties of assessing clinical outcomes, there have been studies that have reported clinical outcomes. Lugtenberg, Burgers, and Westert (2009) conducted a systematic review of 20 studies that evaluated the effectiveness of evidence-based clinical guidelines in the Netherlands. Results showed that the evidence-based guidelines improved the process and structure of care provided to medical patients. However, there was not clear evidence that these guidelines improved patient outcomes. Of the nine

studies that looked at patient outcomes, only six demonstrated small, positive outcomes on some measures of patient health. Three studies showed no impact on patient outcomes. The authors noted that time delays and confounding of outcomes made the process of measuring patient outcomes complex and difficult to interpret.

In the field of social work, numerous studies have shown positive client outcomes when practitioners utilize interventions that have been supported by research (see Campbell Collaboration's systematic reviews for social work; available at <http://www.campbellcollaboration.org>). However, no studies could be located that directly link the training of practitioners in the process of EBP to outcomes of their future clients. Again, as both the Evidence-Based Medical Working Group (1992) and Hatala and Guyatt (2002) discussed, the evidence of the potential link between training and patient outcomes is difficult to collect due to design and methodological challenges. Although there is little research on the clinical outcomes of training on EBP, there is considerably more research on the educational outcomes.

*Educational outcomes.* The educational outcomes on EBP training have been measured by evaluating one or more of these four constructs: (a) knowledge, (b) skills, (c) attitudes, and (d) behavior. I discuss five systematic reviews that examined evidence of all four constructs, and then discuss individual studies that looked at specific constructs.

*Systematic reviews.* Coomarasamy and Khan (2003) and Khan and Coomarasamy (2006) reported on systematic reviews of post-graduate training in evidence-based medical practice. Both reviews found that stand-alone teaching interventions improved EBP knowledge, but not skills, attitudes, or behaviors of medical practitioners. In



contrast, clinically-integrated EBP teaching was found to improve knowledge, skills, attitudes, and behaviors. Khan and Coomarasamy concluded that the teaching of evidence-based medicine should be moved from classrooms to clinical practice to achieve improvements in educational outcomes.

Flores-Mateo and Argimon (2007) also reported on a systematic review of EBP training in postgraduate healthcare education. Overall, small improvements were seen in knowledge, skills, attitudes, and behavior towards EBP. Of the 22 instruments used to evaluate EBP education in 24 studies, only 2 were used in more than one study and only 10 provided at least 2 types of validity and reliability evidence. The authors recommended that future research give greater attention to providing validity and reliability evidence for existing instruments. Other research has shown, in particular, that the evaluation of EBP attitudes and behaviors lags behind the evaluation of EBP knowledge and skills (Shaneyfelt et al., 2006).

Dizon, Grimmer-Sommers and Kumar (2012) reported on a systematic review of evidence-based practice training of allied health professionals (health professionals other than medical doctors and nurses). The researchers included six studies in their review and concluded that EBP training was effective in increasing the skills and attitudes toward EBP. They were unable to determine if the increase in skills carried over to actual practice, and they were unable to differentiate the effectiveness of specific training components on learning outcomes.

Gira, Kessler and Poertner (2004) reviewed twelve meta-analyses on interventions to increase healthcare professionals' use of EBP. Of the following interventions—dissemination of educational materials, continuing education, educational outreach visits,

use of local opinion leaders, audit and feedback, continuous quality improvement programs, use of computers, and mass media campaigns—there was no one intervention that showed success across studies. In general, they found that a combination of intervention methods was more successful than individual methods. In addition, use of printed educational materials, local opinion leaders and continuous quality improvement were classified as weak interventions whereas continuing education and use of computers showed moderately positive effects.

*Individual studies.* Cheng (2003) reported on a double-blind randomized controlled trial designed to evaluate an educational workshop to improve the information seeking behavior of hospital clinicians ( $N = 800$ ). The workshop provided didactic and hands-on training on question formulation and online searching of databases. The training was effective in improving information-seeking skills, knowledge, attitudes and satisfaction with search outcomes. However, the improvement in knowledge and skills eroded by the twelve-month follow-up. The author concluded that additional follow-up measures would be needed to maintain the increase in skills. The next two studies that I review include follow-up measures.

McCluskey and Lovarini (2005) studied the effects of the combination of a two-day EBP workshop and eight months of follow-up support (email, phone calls, on-site visits) on occupational therapists' knowledge, skills, attitudes, and behaviors regarding EBP. They found significant gains in EBP knowledge and skills after the workshop and that these gains were maintained at the eight-month follow-up. However, the EBP behavior of the clinicians did not show meaningful, positive change. During the eight-months following the workshop, over 60% of participants did not read research literature

and over 80% did not appraise research. The researchers concluded that the training and follow-up period were not sufficient to change the practices of the occupational therapists.

Parrish and Rubin (2011) reported on a seven-hour continuing education EBP training for social work practitioners. A local practitioner who was considered to be an opinion leader began the training with an introduction to the need for EBP. Then the workshop provided a combination of didactic instruction on the five steps of EBP and opportunities for participants to practice and receive feedback on the steps. Researchers assessed effectiveness of the training through pre, post, and three-month follow up assessments that used the Evidence-based Practice Process Assessment Scale (EBPAS). The EBPAS has six subscales: familiarity, attitudes, feasibility, intentions, behavior, and knowledge. Researchers found an increase for all subscales that was sustained at follow-up (the behavior subscale was only given at pretest and follow-up as insufficient time had passed between pre and posttesting).

**Theoretical and contextual considerations for training on EBP.** In addition to empirical studies on EBP training, researchers have also developed theoretical frameworks and recommendations for future work in the area of training practitioners to engage in the process of EBP. Michie, Johnston, Abraham, Parker, and Walker (2005), through consensus of interdisciplinary representatives, developed a theoretical framework of constructs that impact implementation of EBP. Created to improve research on effective implementation of EBP training, the framework is composed of twelve domains: (1) knowledge, (2) skills, (3) social/ professional role and identity, (4) beliefs about capabilities, (5) beliefs about consequences, (6) motivation and goals, (7) memory,

attention and decision processes, (8) environmental context and resources, (9) social influences, (10) emotion regulation, (11) behavioral regulation, and (12) nature of the behavior. Michie et al. recommended that their explanatory domain list be used to improve the effectiveness of EBP implementation efforts.

In addition to the constructs identified by Michie et al. (2005), Blisker (2000) reported that professional value, identity, and philosophical approach were constructs that influenced practitioner learning of EBP skills. Gotham (2006) used a case example to highlight the need to attend to multiple levels of structure in EBP training. In other words, EBP educators ought to consider the individual, the organization, and the external environment when engaging in EBP training. Finally, Mullen, Bledsoe and Bellamy (2008) also highlighted the importance of separating research on implementation of an intervention from research on the intervention itself. They also differentiated implementation from dissemination: dissemination is the targeted distribution of information and implementation is the use of strategies to introduce interventions in specific environments. They provided examples of five distinct strategies for dissemination and implementation of EBP: (a) the teaching model, (b) direct implementation of ESIs, (c) a model combining evidence and stakeholder consensus, (d) combining staff training and organizational development, and (e) development of professional infrastructure as agent.

**Challenges to EBP in social work.** The difficulties and challenges of preparing social work practitioners to engage in EBP have been discussed by many social work researchers (Adams, Matto, & Lecroy, 2009; Edmund, Meglvern, & Howard, 2006; Mullen, Shlonsky, Bledsoe & Bellamny, 2005; Raines, 2004; Shlonsky & Gibbs, 2004).

Adams, Matto and Lecroy (2009) produced a comprehensive, and detailed discussion of persistent, if not intractable, challenges to teaching EBP in the field of social work. First, they noted that the problems of social work practice are complex, interconnected, and context specific. Adams et al. challenged the EBP model on the philosophical grounds that it may not be appropriate to the context-specific problems of social work. Next, the authors argued that publication and methodological biases limit the research evidence that is available for use by practitioners. Some types of interventions may be easier to study or may receive more funding. As a result, the process of EBP necessarily privileges certain types of knowledge/interventions to the exclusion of other, perhaps equally efficacious, interventions. Furthermore, EBP training often involves instruction on a hierarchy of research evidence, with randomized-control trials (RCTs) at the top of the hierarchy. However, because of design issues, RCTs are often less generalizable to the work done by practitioners in the field. Thus, Adams et al. questioned the usefulness of the hierarchy. In contrast, Shlonsky and Gibbs (2004) discussed the dangers of eliminating the research hierarchy. They cautioned that EBP could become a catchphrase for any intervention that has research support, no matter the quality of the research. In addition to the question of the utility of the research hierarchy, Adams et al. also questioned how much undergraduate or master's level social workers can be expected to learn about research methodology. Is it possible for social workers with limited methodological training to interpret research evidence with the appropriate level of nuance needed to apply research findings to practice?

Adams et al. (2009) also suggested that the role of theory in EBP has not been adequately explicated. Without theoretical grounding, Adams et al. speculated that

clinicians might become mere technicians who implement prescribed practices rather than independent professional practitioners. In addition, Adams et al. questioned the extent to which social workers should engage in practices that are supported by research but for which they themselves have limited training. Finally, they questioned how social workers balance their clinical wisdom with research findings. No clear guidelines exist for such a balance.

In addition to the concerns addressed by Adams et al. (2009), time constraints, practitioner resistance, lack of access to research, and lack of relevant research have all been listed as other challenges to training in EBP (Edmund, Meglvern, & Howard, 2006; Mullen, Shlonsky, Bledsoe & Bellamny, 2005). Finally, Rubin (2007) suggested that the most significant challenge in teaching EBP is the lack of support for the EBP process in the context of fieldwork training and supervision. Because clinically-integrated teaching of EBP has been shown to be most effective for EBP training of health professionals (Coomarasamy & Khan 2003; Khan and Coomarasamy 2006), the clinical-integration of EBP training in social work is of special interest. In the next section of this literature review I discuss research on integrating EBP into the context of field placements for social work students.

**Social work field placements and training on EBP.** The partnerships between universities and agencies that host fieldwork practica have been understood to be key in the advancement of EBP in the field of social work (Franklin, 2007; Mullen, Bellamy, Bledsoe, & Francios, 2007; Proctor, 2007; Rubin, 2007). Edmond, Meglvern, Williams, Rochman, and Howard (2006) sent out a 25-item questionnaire to fieldwork supervisors ( $N = 235$ ) who worked with a university that had recently adopted EBP as the guiding

principle of its curricula. The researchers found that the majority of field instructors believed that EBP was a useful practice idea. However, less than half reported that they engaged consistently in EBP. Field instructors indicated that they were more likely to rely upon experience, intuition, and colleague advice rather than finding research on interventions that had been empirically tested. Researchers concluded that the school of social work should find ways to make social work research more accessible to fieldwork supervisors and provide EBP training for supervisors.

Mullen, Bellamy, Bledsoe, and Francios (2007) reported on a pilot project (Bringing Evidence to Social Work Training—BEST) that fostered a partnership between Columbia University School of Social Work and three social work agencies for field training of EBP. One of the greatest partnership challenges was that the culture of agencies that hosted fieldwork practica was one where knowledge was assumed to be stable. These agencies did not possess the structures or motivation to engage in EBP. Based upon their pilot work, the researchers recommended that the process of EBP be carried out in the context of a team, and that practitioners read research summaries (or other publications designed for practitioners) rather than individual research articles. The researchers also recommended incorporating EBP training into existing agency trainings and meetings. Now that I have reviewed a broad base of literature on EBP, I next review literature that is specific to applying research to practice in the context of education.

### **Linking Research to Practice in Teaching and Teacher Preparation**

To date, the field of education has not clearly defined the process in which teachers engage to apply education research to teaching practice. As a result, the broad construct of linking research to practice has been addressed in disparate ways in the

literature. Relevant terms used to refer to the construct of applying research to practice include but are not limited to: *research utilization*, *application of research to practice*, *research implementation*, *scientifically based practices*, *evidence-based practices*, *evidence-informed practices*, *research to practice gap*, and the *research to practice divide*. The variety of phrases used to refer to this construct reflects both the importance of the construct and a lack of consensus in the field of education on a systematic way to discuss and research this construct.

**Linking research to practice with inservice teachers.** I begin this section by reviewing studies of *inservice* teachers who have linked education research to teaching practices. The purpose of this section is to identify the skills that are utilized when teachers link research to practice. This information can be used to suggest key areas on which to focus in the preparation of preservice mathematics teachers. I have organized relevant articles around three central themes: how teachers understand research, how teachers implement research, and the social context in which teachers implement research.

***How do inservice teachers understand research?*** In this section, I present findings from five studies that have contributed to the field's knowledge of how inservice teachers understand education research. Two studies reported findings on how teachers read research. One study examined how teachers perceive the format of the communication of education research. Another study examined how teachers view education research as part of a comprehensive professional development program. Finally, the fifth study linked prior experience of research with a more complex view of



research. These five studies are complementary and, taken together, help to define the ways in which teachers understand research.

Zeuli (1994) reported results from a qualitative study of 13 inservice teachers. This study examined *how* teachers understood the education research they read. Through the analysis of data from interviews with teachers, the researcher found a distinction between knowledge of participants who read only to understand the product of the research (what could directly be applied to the classroom) and knowledge of those teachers who read to understand the process of the research. Teachers who read only for a product to be applied to the classroom had a shallow, consumerist approach to the research. Zeuli concluded that the shallow reading of the research articles did not contribute to the education of these teachers. On the other hand, some teachers read to understand the process of the research study. These teachers communicated understanding of the study's argument and the importance of the study. They took a more critical stance toward the research, and the researcher concluded that this resulted in an educational benefit for these teachers. Overall, Zeuli highlighted the importance of thinking about *how* teachers read research instead of *if* they read research: teachers who only looked for products in research were much less informed about the research compared with those who pursued understanding of the process of the research.

Kennedy (1997) also contributed to the field's knowledge of how teachers read education research. Kennedy reported on a qualitative analysis of two teachers' responses to assigned research studies. Kennedy found that the teachers understood the articles fairly well, but that they used their own personal experiences and values to judge the

validity of the research. In addition, the teachers used the research that they read to validate their own teaching practice.

Landrum, Cook, Tankersley, and Fitzgerald (2007) expanded the conversation about how teachers understand research by their examination of teacher preference for the way in which research is presented. This quantitative, experimental design study found that teachers ( $n=127$ ) rated education research as more useable when it was presented in a personal, narrative format versus a data-based format (scholarly, empirical article). In addition, greater years of teaching were associated with giving lower usability ratings to the research.

De Geest (2010) investigated how teachers understand education research by examining how they perceived research in their professional development experiences. Through the use of grounded theory, De Geest analyzed online questionnaires, field notes of discussions, and observations to examine the effects of including education research in 30 ongoing professional development programs. Teachers reported that the inclusion of education research in professional development stimulated their thinking about issues related to teaching and learning, fostered their professional growth, gave them confidence that the professional development was worth implementing, and added credibility to the professional development. These findings suggest that teachers understand education research both to be pertinent to professional development and to have a positive effect on their growth. A significant limitation of this study is its lack of discussion on sampling. The study did not report a systematic method of analyzing teacher perceptions. Therefore, this study can serve as a proof of concept; there are some teachers who understand education research as having a key role in professional development, but the findings do

not support the conclusion that this is how most teachers understand the role of education research in professional development.

Finally, Ratcliffe et al. (2005) reported on how science educators think about education research. This study was designed so that the thoughts of teachers who had a history of participating in education research ( $n=20$ ) could be compared to the thoughts of those teachers who had no prior history of education research ( $n=21$ ). Data for this study were collected through interviewing and focus groups. Researchers found that teachers with prior history of education research had more expansive definitions of research. Furthermore, these teachers were more likely than teachers without research experience to see the results of education research as having a direct impact upon their teaching. The teachers with histories of participating in education research communicated a greater willingness to incorporate ideas from research into their own practices and a greater ability to engage in reflection about their practices. The authors concluded that teachers with research experience were more likely to view their teaching practices through an evidence-informed lens.

***How do inservice teachers implement education research?*** In this section, I present findings from empirical articles that have examined how inservice teachers implement education research. The studies provide insights about the difficulties that teachers encounter when trying to implement research in the classroom setting.

Antil, Jenkins, Wayne, and Vadasy (1998) reported on the implementation of cooperative learning in elementary school classrooms. In this study, 93% of teachers ( $n=84$ ) reported on a written survey that they regularly used the research-supported practice of cooperative learning in their classrooms. However, when a subset of these

teachers (n=21) was interviewed, researchers determined that, at most, 24% were using cooperative learning in a manner consistent with the evidence supporting this practice. This study is significant because it demonstrated that it can be difficult for teachers to evaluate their own implementation of a research-based practice. Teachers thought they were applying a researched strategy to their classroom teaching, but, in fact, they modified that strategy in ways that deviated from its intent as described in the literature.

Klinger, Ahwee, Pilonieta, and Menendez (2003) documented a rigorous, qualitative study of 29 teachers who implemented research in their teaching practices. For two weeks in the summer, these teachers participated in an institute where they learned about research on effective strategies for reading instruction. These teachers then received extensive follow-up support throughout the year. Researchers collected data about research implementation in the form of interviews, teacher logs, researcher logs, classroom observations, and implementation validity checklists. Data analysis showed that, of the 29 teachers, 9 were classified as *high implementers*, 9 were *moderate implementers*, and 11 were *low implementers*. Despite extensive, weekly support for implementation, 25% of the teachers were low implementers—teachers who implemented the practices very little or not at all. In addition, while treatment fidelity was fairly high among the high implementers, it was low with the moderate and low implementers. These data are consistent with the low fidelity reported in Antil et al. (1998). Specific barriers to implementation were reported as lack of instructional time, lack of administration support, and student behavior. Specific facilitators for implementation included student appreciation of the strategies, improvement in student learning, administrative support, and adequate preparation in the research strategy.

Vanderlinde and van Braak (2010) provided evidence on the barriers and facilitators of implementation. They reported results from four focus groups of inservice teachers ( $n=12$  per group) and a questionnaire completed by inservice teachers ( $n=68$ ). Researchers identified the following facilitators and barriers for the teachers who were asked about applying research to practice. Barriers were ambiguity, lack of applicability of research findings, technical use of language, and a need for opportunity to link their understanding of the research to their knowledge of teaching. A significant limitation to this study is that it asked for teachers' opinions on implementation, but the study did not report on the actual implementation practices of teachers.

***What are the social contexts in which inservice teachers link research to practice?*** The studies discussed in this section provide information about the influence of the social context on a teacher's ability to link research to practice. Specifically these articles address top-down leadership environments, the complexity of the professional relationships that influence implementation, and the context of sustained, systematic change.

*Top-down leadership approaches to research implementation.* Lamb, Cooper, and Warren (2007) documented the conflicts that arose when teachers ( $n=10$ ) were mandated by their principals to participate in a program that involved the implementation of mathematics education research in their algebra classrooms. These teachers did not want to understand why new, experimental techniques might work or not work. Instead, they wanted to be given practical guidelines as to what they should do in the classroom. However, they were unable to find time to implement the guidelines that they were given. Furthermore, their content knowledge was low and this hindered their ability to engage

intellectually with the research projects that were taking place in their classrooms. Lamb et al. identified numerous conflicts that were experienced between the teachers and the researchers in this project. Because of these conflicts, the mandated teacher participation in the implementation of research-based strategies was largely unsuccessful.

Schmidt (2011) reported on a large, multi-state grant designed to improve student achievement in mathematics and science in 61 districts in the Midwest. This project can be considered top-down in that individual teachers and principals did not choose to be a part of the program but were mandated to do so by their districts. Schmidt noted several difficulties that were encountered in this program. It is important to point out that these observations are not findings per se but observations relevant to the context in which teachers were required to implement mathematics education research. Schmidt noted that the teachers were unable to draw implications from the researchers' presentations of the data; teachers needed to be told explicitly what the implications for classroom teaching were.

Civil (2011) reported on researchers' attempts to communicate with teachers the results of a study which demonstrated that tracking had harmful effects on students' mathematics achievement. Civil reported that teachers did not want to accept the research results because of their low opinion of the end-of-course assessment used to evaluate student performance and because of the teachers' own experiences related to students' behaviors. Although anecdotal, these observations by Civil provide insights about decisions that teachers make with regard to the implementation of education research in their classrooms. Because the teachers did not agree with the research finding that tracking had a negative effect, teachers wanted to continue their previous tracking

practices. In summary, the three examples cited here highlight the potential pitfalls and limitations to the use of top-down leadership approaches to research implementation. The next section will examine the relationships that are involved when teachers implement education research in their classrooms.

*Professional relationships that influence research implementation.* The studies discussed in this section explore the role that relationships between teachers and researchers play in research implementation. Potari, Sakondis, Chatzigoula, and Manaridis (2010) provided results from a qualitative study of the relationships formed over a four-year collaboration between classroom mathematics teachers and mathematics education researchers. The relationships between teachers and researchers were complex and evolved over time. From the perspective of the education researchers, the teachers learned to form a community of inquiry through critical reflection on their teaching practices. This process was not easy and involved much negotiation and flexibility on the part of both the researchers and the teachers. Initially, teachers and researchers had different agendas and different ways of discussing their agendas. Over time, teachers and researchers both learned to form a common community of inquiry through the process of making the teaching act public and thinking of global rather than local teaching issues.

Martin, Strutchens, Stuckwisch, and Qazi (2011) reported on a network of university-school partnerships that were created to improve K-12 mathematics achievement in eastern Alabama. Although the focus of this project was on student achievement and not on the relationships between researchers and teachers, the mathematics education researchers found the relationships essential to the ultimate success of the project. They stated:

We have learned, too, not to underestimate teachers. Although some of the partners questioned whether teachers would engage in reading mathematics education research, we have found that teachers will respond in positive ways to professional development that is based on national and state standards and contains best practices. Finding teachers who believe in the project and have implemented the curriculum to serve as presenters is also essential to the project's success. They increase credibility among participants and encourage continued participation. (p. 116)

This statement speaks to the importance of the full participation of teachers in the implementation of research aimed at improving student learning. The final two studies in this section provide clarity as to how researchers and teachers might achieve the strength of relationship necessary to maximize success of the implementation of new research in the classroom.

Anagnostopoulos, Smith, and Basmadjian (2007) provided a theoretical framework and precise vocabulary to discuss the teacher educator (researcher) and teacher (practitioner) relationship. The ability of the teacher educator to communicate with the teacher and vice versa was referred to as *horizontal expertise*. This expertise both contributes to and is the result of the creation of *boundary objects*. Boundary objects are material products that can facilitate connection, communication, and coordinated work between diverse groups—such as teachers and teacher educators or researchers. After providing a theoretical framework and vocabulary to enhance the discussion, Anagnostopoulos et al. reported on the creation of a rubric that inservice teachers used to mentor new teachers. This rubric was a boundary object because it was developed by



both teacher educators and inservice teachers. The creation of the rubric was full of challenges, and yet the final product was successfully used to bridge the traditional gap between the research work of the university researchers and the teaching practices of inservice teachers.

Sinnema, Sewell, and Milligan (2011) described a project in which 26 teachers engaged in collaborative inquiry with 6 researchers over the course of a year. The teachers were successful both in finding education research that was relevant to their teaching concerns and in applying that research in the context of inquiry-based teaching. Two vignettes, or cases, were provided as illustrations of the successful application of research to practice by the classroom teachers.

Unlike Anagnostopoulos et al. (2007), Sinnema et al. (2011) did not discuss the quality of the collaborative relationships between teachers and researchers. Given the challenges to classroom implementation that have been addressed thus far, it would have been helpful if information were provided about the relationships between the teachers and researchers as well as a quantification of how well the other teachers in the study (the ones not represented by the cases) were able to apply research to practice. Although this evidence was not provided, the authors did provide anecdotal evidence of the positive impact that this evidence-informed inquiry had on student outcomes—particularly on student groups who had a history of inequitable access to education. Through interviewing and observations, teachers recorded increases in student interest and knowledge after research implementation. Although no causal link could be established between the research implementation and increases in student learning, there was an association between the two. The link between implementation and student outcomes is

critical to the field's evaluation of linking research to practice for the achievement of equity in classroom teaching. The two cases discussed in Sinnema et al. (2011) serve as a proof of concept. It is possible for teachers, in the context of collaborative relationships with researchers, to find, understand, and apply education research to their classroom teaching. Moreover, this linking of research to practice was associated with improved outcomes for students. Additional research is needed that provides a qualitative analysis of what was working in the teacher researcher relationships as well as a quantitative analysis of what this experience is like for *all* the teachers who are involved. Furthermore, the link between research implementation and student outcomes needs to be examined more closely.

*Systematic change.* In this section, I discuss two studies that addressed the need to link research to practice from comprehensive, systematic perspectives. Miller, George, and Fogg (2005) reported on a school-wide effort to implement specific research-based practices aimed at supporting the achievement of their student population. Within the context of an alternative day school for students with severe emotional and behavioral disorders, Miller et al. described the multiple factors needed to achieve system-wide change. The school's comprehensive approach to systematic change required the organizational structure of the school to be enhanced to support change. Teachers, who were asked to implement research-based practices, were not only supported by administration, but they were also viewed as part of the school's problem-solving capacity and were directly engaged in the work of mediating barriers to implementation. Miller et al. demonstrated the possibility of sustained teacher implementation of research-based teaching practices within the context of school-wide commitment to change.

Fuchs and Fuchs (2001) reported on a large school district's sustained use of a research-based strategy, Math Peer Assisted Learning Strategies (Math PALS). This case study provided evidence of improvement in student achievement through the use of this research-supported strategy. Based on the results of this study, the authors identified five principles for sustaining teacher use of research-based practices in this school district: (1) There was a key individual who advocated for the change; (2) The schools implementing the research had at least some control over their resources and could reallocate these to support change; (3) The implementation of the research-based strategy occurred in the context of accountability for student outcomes; (4) Teachers had tolerance for initial problems with the implementation; and (5) Successful implementation was recognized and praised publicly.

**Linking research to practice with preservice teachers.** I have organized the literature on linking research to practice in teacher preparation around the following themes: the meaning of education research, experiential learning, systematic approaches to linking research to practice, and longitudinal effects.

*The meaning of education research.* Gitlin, Barlow, Burbanks, Kauchak, and Stevens (1999) distributed a questionnaire and conducted interviews at the beginning and end of an inquiry-orientated teacher preparation program to determine how preservice teachers ( $n=37$ ) thought about research. Overall, preservice teachers viewed education research as a technical endeavor that was undertaken solely for the purpose of discovering specific methods that could improve teaching. Preservice teachers expressed concern that researchers interpreted findings in ways that supported predetermined perspectives and that results could not be applied to typical classroom settings. Preservice

teachers also preferred the input of experienced teachers over the use of education research for guidance on their teaching. To address preservice teachers' devaluation of research, the authors suggested that teacher preparation programs might provide opportunities for preservice teachers to examine critically the goals of education research.

Gore and Gitlin (2004) also reported results on how inservice ( $n=147$ ) and preservice ( $n=85$ ) teachers viewed education research. Although approximately three-fourths of *preservice* teachers thought that education research addressed their concerns about teaching at least some of the time, less than one-tenth of *inservice* teachers thought research addressed their concerns about teaching. Both preservice and inservice teachers expressed concern that the results of education studies would not apply to their specific classrooms. In contrast to Gore and Gitlin, Greenwood and Mabeady (2001) reported that 88% of preservice teachers ( $n=111$ ) were unaware that a gap existed in their profession between research and practice.

***Experiential learning about applying research to practice.*** In this section I discuss studies that address experiential learning in the following forms: instructor modeling, embedded design, methods coursework linked to practicum experience, research coursework, qualitative research experience and implementing research in the elementary school classroom.

Allinder (2001) examined preservice teachers' ( $n=42$ ) preference for research-based teaching strategies that had been modeled by their instructors. Pre-service teachers indicated that they had more knowledge of modeled strategies and that they believed these strategies to be more helpful for students than strategies about which they had read but which were not modeled by their instructors. The results from this study provided

evidence of preservice teachers' preferences for learning about research-based strategies through instructor modeling, but no evidence was given to show that modeling equips preservice teachers to apply these strategies to their future teaching practices.

Bain, Lancaster, Zundans, and Parkes (2009) engaged preservice teachers in experiential learning about research-based practices through the use of embedded design. Embedded design requires building a practice into multiple levels of an educational experience. Two evidence-based practices (cooperative learning and peer-assisted learning) were introduced in a teacher preparation course at the following levels: the academic knowledge level, an active experience level, a real-world application with feedback level, and a personal impact (course assessment) level. The results were ambiguous as to whether learning outcomes were improved due to exposure to embedded design.

McDonnough and Matkins (2010) examined whether differences in practicum supervision affected the ability of preservice teachers to connect research to practice. They compared two types of practicum experience in preservice elementary teacher preparation for science teaching. At one institution, preservice teachers ( $n = 97$ ) took a science methods course and a separate teaching practicum. At a second institution, preservice teachers ( $n = 44$ ) took a science methods course and a concurrent teaching practicum that was supervised by the instructor of the methods course. McDonnough and Matkins found that preservice teachers who took the practicum course supervised by the methods instructor reported increases in their self-efficacy beliefs about teaching science. However, the difference between the two groups in self-efficacy scores was not statistically significant. The researchers also used qualitative data analysis to conclude

that teachers with the same instructor and supervisor were better able to connect research to practice. However, sufficient information (i.e., methodology, inter-rater reliability) was not provided to permit the reader to determine the robustness of the results.

Harrison, Dunn, and Coombe (2006) reported on the incorporation of a research methods course in a preparation program for early childhood teachers ( $n=70$ ). The course included the following experiences: lectures, readings, small group work, critiquing and discussing research articles, journal writing, conducting a small research project, and mounting a conference to present their research. Harrison et al. found that preservice teachers grew significantly in their understanding of education research. Although preservice teachers reported that they neither enjoyed reading education research nor that they understood research well, they reported that reading research reports was indispensable to their applied research project. Indeed, preservice teachers indicated that reading education research was of no use to them without its practical application in their research projects. The researchers concluded that preservice teachers gained an appreciation for the fact that teaching practice can be improved by applying the findings of relevant research to practice.

Del Carlo, Hinkhouse, and Isbell (2010) described the connections between preservice teacher experience with qualitative research (outside of a K-12 classroom) and the development of reflective practices. Building on information about undergraduate research gleaned from fields other than education (e.g., biology, physics, chemistry), these researchers provided undergraduate preservice teachers with the opportunity to conduct qualitative research on questions not related to teaching practices. They provided an illustration of this connection through two cases of preservice teachers who

were given the opportunity to engage in qualitative research. The two cases provided anecdotal evidence that the experience conducting qualitative research improved their ability to reflect critically.

Everett, Luera, and Otto (2008) explored two research questions: Can preservice teachers apply education research to their classroom teaching, and can they relate their teaching experiences back to published research? Preservice teachers ( $N=144$  from 9 semesters) were required to conduct an action research project as a part of a capstone course—the sixth course in a sequence of science content and methods courses. The action research project required preservice teachers to search primary literature to learn about common student misconceptions about the science topics that they would teach. Then, preservice teachers conducted a pre-assessment to determine the needs of their students. Following the preassessment, preservice teachers had to develop an intervention, backed by education research, that would be responsive to the documented needs. A post-assessment was conducted, and preservice teachers were required to summarize their action research projects in the form of written reports. These reports provided the data for the qualitative research conducted in this study. The researchers utilized grounded theory in their data analysis. They found that 100% of preservice teachers were able to find published education research related to their teaching topics and 96% were able to use this information to inform their preassessments. In contrast, only 16% of preservice teachers were able to link their results back to the published research base. The researchers conducted a follow-up study that provided scaffolding for preservice teachers on how to link their teaching results back to the research base.

Although this follow-up study was much smaller ( $n=17$ ), 94% of preservice teachers were able to link their teaching back to the research base.

*Longitudinal studies.* Cady, Meier, and Lubinski (2006) reported on a follow-up study six years after an original study (Lubinski, Otto, & Rich, 1996) examined the effects of experience with Cognitively Guided Instruction (CGI) on preservice teachers' application of research to practice. In the original study preservice teachers were placed in classrooms with teachers who utilized CGI. Preservice teachers learned through classroom observation experiences and formal discussions about CGI practices. At the end of their final year in teacher preparation, these preservice teachers demonstrated changes in beliefs about teaching and learning that were consistent with CGI, but they failed to demonstrate an ability to implement corresponding changes in their teaching practices. However, six years later, Cady et al. found that these former preservice teachers were making teaching decisions consistent with CGI practices. This is an important finding because it demonstrates the long-term effect of preservice teacher experiences with evidence-based practices (in this case CGI).

Umbeck (2011) provided a first-hand account of a novice teacher who documented her struggle to implement the teaching practices that she valued in her preservice preparation. This narrative account illustrated how a new teacher can fail to implement practices that she believes to be valuable to student learning and then slowly start incorporating those practices over time. This article is a first-hand account of the phenomenon that was identified in the Cady et al. (2006) article—namely that a preservice teacher may gain understanding of how to apply research to practice during preservice education, but may not actually be able to implement this understanding until



several years into the teaching profession. This concludes the review on linking research to practice in the context of teaching and teacher preparation. Because the process of linking research to practice can be conceptualized as a routine of practice, I discuss literature on this topic in the following section.

### **Routines of Practice**

The preparation of teachers is a complex task, and there has been lack of clarity and consensus among those who have responsibility for teacher preparation as to the curriculum and foci of preparation programs (Morris & Hiebert, 2009). Currently, there are several groups within the field of mathematics teacher preparation who have been working to define what can be thought of as the central practices of mathematics teaching (Ball, Sleep, Boerst, & Bass, 2009; Boerst, Sleep, Ball, & Bass, 2011; Franke, Grossman, Hatch, Richert, & Schultz, 2006; Kazemi, Lampert, & Ghouseini, 2007). Once these central practices have been identified, the intention is for them to be incorporated into teacher preparation programs. This type of work has been conceptualized by Grossman et al. (2009) as a pedagogy of professional preparation that deconstructs complex practices and incorporates pedagogies of enactment that allow preservice teachers to engage in the core teaching practices at varying levels of approximation.

Ball, Sleep, Boerst, and Bass (2009) called these core practices *high-leverage* practices and provided the following list of defining criteria for the identification of these practices specific to the field of mathematics education:

Criteria based on examination of the work of mathematics teaching:

1. Supports work that is central to mathematics
2. Helps to improve the learning and achievement of all students

3. Is done frequently when teaching mathematics
4. Applies across different approaches to teaching mathematics.

Criteria necessitated by our teacher education context:

5. Can be articulated and taught
6. Is accessible to learners of teaching
7. Can be revisited in increasingly sophisticated and integrated acts of teaching
8. Is able to be practiced by beginners in their field-based settings. (p. 461).

Because routines of practice are defined by what teachers *do* in addition to what they *know*, it is important to consider how preservice teachers will be able to engage in these practices.

Grossman and colleagues (2009) argued for the importance of *approximations of practice* in the preparation of teachers. Such approximations provide opportunities for the novice teacher to begin engaging in teaching practices under controlled conditions. By allowing novice teachers to engage in approximations of practice, teacher educators can provide strategic support and feedback to the novice such that the novice can develop and strengthen teaching skills and practices. In addition, Lampert (2010) has argued for the importance of allowing teachers to learn in, from, and for practice. Novice teachers must be engaged in practice in order to learn in and from practice. Herbst, Aaron, Bieda, and Moore-Russo (2012) also suggested that cognitive psychology research on active learning provides yet another perspective on why having teachers engage in the practices of teaching can be an important and effective strategy.

In the field of education, an emphasis on routines of practice has its philosophical roots back in the work of John Dewey (1904/1965). Dewey argued that a laboratory approach to preservice preparation would be most effective. Instead of submersing a teacher in a classroom fulltime, a laboratory approach would create focused opportunities to experiment and refine teaching practices. Almost a century later, Ball and Cohen (1999) sparked renewed interest in routines of practice through a chapter that discussed the importance of grounding teacher preparation in the practice of teaching. They suggested that learning from artifacts such as student work or classroom videos could serve as the basis for developing knowledge about teaching and learning.

Grossman and colleagues (2003) explored how different professions prepared novices for professional practice. In a three-year study, they analyzed the professional preparation of clergy, clinical psychologists and teachers. Growing out of this work, Grossman et al. (2009) created a three-part framework to describe how professional practice can be taught. They explicated how the use of *representations*, *decompositions*, and *approximations of practice* can prepare the novice for professional practice:

Representations of practice comprise the different ways that practice is represented in professional education and what these various representations make visible to novices. Decomposition of practice involves breaking down practice into its constituent parts for the purposes of teaching and learning.

Approximations of practice refer to opportunities to engage in practices that are more or less proximal to the practices of a profession. (p. 1)

Moss (2011) recommended adding a fourth concept to the framework. Moss argued that *conceptions of quality* was a critical construct for both the understanding of

teaching practices and the improvement of those practices. Moss defined conceptions of quality as “what educators need to judge whether some instance of practice is more or less mature, sophisticated, or successful, and to offer direction for improvement or development” (p. 2879). Moss further explained that conceptions of quality can be understood in terms of three dimensions: the grain size of practice that is being evaluated, the kinds of criteria or qualities that are being highlighted, and the ways the variations from less to more advanced are represented.

Moss’ recommendation to attend to conceptions of quality points to the need to address, in a systematic fashion, the assessment of routines of practice. Clearly articulating the conceptions of quality that are expected for a given routine of practice will be a critical step in producing effective assessments of routines of practice. However, as of yet, the field of mathematics teacher preparation has not advanced to the point of publishing systems of assessment for the routines of practice that have already been identified.

For example, Boerst et al. (2011) provided an illustration of how the high-leverage practice of leading a mathematical discussion has been incorporated into methods courses in an elementary preservice teacher preparation program at the University of Michigan. They described how preservice teachers learn to engage in this practice of leading a discussion and discussed some of the challenges that must be faced when assessing preservice teachers’ abilities to lead a mathematical discussion. Although Boerst et al. discussed the need for assessment, they did not provide information on their assessment of this practice.

Kazemi, Franke, Lampert (2009) provided a comprehensive list of activities in which a teacher educator must engage when approaching teacher preparation from a *pedagogy of practice*. A pedagogy of practice involves a fundamental stance toward teacher preparation that emphasizes the importance of learning in, from, and for practice. When a teacher educator engages in a pedagogy of practice, that educator takes responsibility for:

- exhibiting, demonstrating, and naming the elements of an instructional activity;
- situating the activity in theoretical and empirical evidence that is likely to result in student learning;
- giving novices the opportunity to deliberately practice the elements of the activity that are “routine” with coaching from teacher educators;
- structuring collaborative work on problems of teaching practice so as to attend to the development of novices’ knowledge of important mathematics and their knowledge about how students make sense of that mathematics in ways that are connected with that work;
- scaffolding novices’ preparation for doing the activity with particular elementary level learners in ways that call attention to important mathematics and students’ ways of making sense;
- rehearsing the enactment of the plans for doing the activity so as to provide deliberate practice of its routine elements as well as opportunities to respond in a principled way to the kind of non-routine information that comes from students;
- organizing opportunities for novices to teach using the activity and to record

their practice and their students' work;

- analyzing with novices how an Instructional Activity can maintain its integrity while playing out differently in different classroom contexts;
- assessing the learning of novices around the key practices that are embedded in the activity;
- refining the design of the Instructional Activity in consideration of what elementary mathematics students are able to learn with it. (p. 15)

To summarize, teacher educators who engage in a pedagogy of practice engage in the practices listed previously so that their students (preservice teachers) will be prepared to learn in, from, and for practice (Lampert, 2010). Having reviewed the literature on routines of practice, I now transition to reviewing the literature related to the performance assessments that can be used to assess routines of practice.

### **Performance Assessments in Teacher Preparation**

In this section, I discuss studies that have examined the role of performance assessments in the context of teacher preparation programs (TPPs). In order to communicate effectively about the strengths and weaknesses of these studies, I first provide a theoretical background to the subject of performance assessments. I define terms and discuss issues of validity and reliability that are specific to performance assessment scores. After building context for the empirical review, I then present findings from studies involving performance assessments in TPPs.

**Defining performance assessments.** In both public and scholarly discourse, there are several terms that are often used interchangeably with the term *performance assessment*. These include *competence based assessment*, *direct assessment*, *alternative*

*assessment, authentic assessment, and innovative assessment.* Messick (1984) made a distinction between the terms *competence* assessment and *performance* assessment. He defined *competence* as what a person knows or does under ideal circumstances. In contrast, *performance* indicates what a person knows or does within a specific context that includes affective, motivational, attentional, and environmental factors. Messick warned that a student's competence might not be revealed in a performance assessment due to contextual factors:

Although competence may be defensibly inferred from correct task performance, especially if consistently demonstrated across related tasks, as a general rule it is dangerous to make inferences about competence, or the lack thereof, from incorrect performance. To do that requires the discounting of a variety of plausible rival sources of poor performance, such as inattention, anxiety, low motivation, fatigue, adverse testing conditions, and insufficient test-wisness.  
(p.227)

It is important to note that Messick's distinction between *performance* and *competence* led to the validity implications that are implicit in the quotation cited above. Researchers and test users who understand that performance assessments are context dependent will interpret TPP participants' performances in light of all contextual factors and will refrain from making unsupported claims about competence when performances provide insufficient evidence to make such claims.

In contrast to Messick's distinction between competence and performance, Jonsson and Mattsson (2011) defined competency as being able to "act knowledgably in relevant situations . . . where competency depends not only on the individual but on

contextual factors and the actions of others” (p. 170). The fact that Jonsson and Mattsson defined competence in a manner that is incompatible with Messick’s definition illustrates the potential for confusion and misinterpretation that is based solely on the implicit understanding of foundational concepts in a system of assessment. Each and every construct, beginning with the most simple and fundamental, must be defined in detail in performance assessments—beginning with the term *performance*. Anticipating the empirical review of performance assessments in TPPs that follows, one can see that the definition of *performance* could have significant impact on how researchers or other stakeholders interpret study findings.

**Validity and performance assessments.** The 1999 AERA, APA, and NCME *Standards for Educational and Psychological Testing* (referred to as the *Standards* throughout the rest of this document) defines validity as “the degree to which evidence and theory support the interpretations of test scores entailed by proposed uses of tests” (p. 9). This statement makes clear that the validity of assessment scores is directly tied to the purpose(s) of the assessment. Snyder (2009) highlighted the fact that often the following *three* purposes are intended for performance assessments in the context of teacher preparation: individual candidate evaluation (licensure), accountability for teacher preparation programs, and formative assessment for teacher candidates. Snyder suggested that the needs for these three functions must be balanced against each other. For example, tailoring a performance assessment to make scores more reliable for high stakes purposes may mean that the scores provide less information to support teacher professional development. Snyder’s suggestion of balance stems from his acknowledgement that having three separate, comprehensive performance tests may not



be feasible in teacher preparation programs. However, best measurement practice, as stated in the *Standards* (AERA, APA, NCME, 1999), calls for evidence of validity and reliability for each intended use of a test. This is a more demanding, if less practical, standard than is suggested by Snyder's call for balance.

Cronbach (1969) stated that the "responsibility for valid use of a test rests on the person who interprets it" (p. 51). This statement, which is also reflected in the *Standards*, places the onus for demonstrating validity of scores squarely on the shoulders of the test user. In the case of performance assessments in TPPs, the user, whether that is the federal government, state government, or individual TPP, cannot rely on the test developer's validation studies if the user intends to utilize the test for a purpose other than intended by the developer. For example, a high-stakes performance assessment might be shown to be valid for individual evaluation of teacher candidates. However, problems can arise when that same performance assessment is used for teacher preparation program accountability. In this case, fieldwork experiences can confound data interpretation. If the field experiences do not line up with the vision of effective teacher of either the assessment or the teacher preparation program, then the program will have a difficult time interpreting the data from their graduates' performance assessments and transforming that interpretation into action that can improve the program. In the next section of this review, I discuss validity frameworks that can be used for judgments about validity after the purpose of an assessment has been established.

Messick (1980, 1981, 1995) advanced a unified view of assessment validity that requires an overall judgment about the interpretation, use, and consequential aspects of test scores. He argued against the idea that there are many different types of validity and

against the notion that researchers can simply choose one type of validity evidence among a list of potential options. Linn, Baker, and Dunbar (1991) were also concerned about narrow conceptions of validity, particularly related to performance assessments. They provided the following criteria as a means of expanding the understanding of validity in the context of performance assessments: consequences, fairness, transfer and generalizability, cognitive complexity, content quality, content coverage, meaningfulness, and cost and efficiency. Baartman, Bastiaens, Kirschner, and Van der Vleuten (2007) also proposed a framework for evaluating competence-based assessments with similar criteria: authenticity, cognitive complexity, fairness, meaningfulness, directness, transparency, educational consequences, reproducibility of decisions, comparability, and cost and efficiency. Fundamentally, Messick, Linn et al., and Baartmann et al. share similar views of validity with regards to performance assessment. A particular researcher may find one framework more efficient than another, yet each framework advances an expansive view of validity that ought to be kept in mind when evaluating studies involving performance assessments.

In addition to the validity criteria listed in the above frameworks, Baker (2008) has argued for the importance of including instructional sensitivity as a validity criterion when an assessment is given in the context of educational accountability. Given the fact that TPPs are likely to use their own performance assessments as a means of program development and that state and federal governments plan on using performance assessments as a means of holding TPPs accountable for the quality of teaching candidates, it seems reasonable to highlight this validity criterion. Baker argued that if the measure is not sensitive to instruction, then it is absurd to use the measure for

instructional accountability. Based on her own research and extensive literature reviews, Baker offered the following steps to providing validity evidence of instructional sensitivity:

- Analysis of the test to determine whether content and skills are well described, appropriately transparent, with bounded domains
- Deep sampling for fewer standards
- Subtests or tasks addressed to components of outcomes (prerequisites)
- Transfer tasks to assure that test practice is not exclusively the method of choice
- Better, scalable and quick turnaround measures of classroom practice
- Instructional options for underserved populations (p. 11)

Baker argued that assessments would need specificity (deep sampling) in order to have instructional sensitivity. However, although assessment specificity can increase validity, it can also reduce reliability (e.g., specific sampling could lead to multi-dimensionality or a reduction in inter-rater reliability). The trade-off between actions that increase validity and decrease reliability should be kept in mind when reading the next section on reliability. Overall, the conflict between conditions that increase validity and those that increase reliability helps to expose the messiness of what is sometimes referred to as an “objective” assessment score.

**Reliability and performance assessments.** AERA, APA, and NCME (1999) defined reliability as “the consistency of measurements when the testing procedure is repeated” (p. 25). Although it is common for researchers to report isolated measures of reliability, such as the Cronbach alpha, there exists a powerful framework, generalizability theory, which allows researchers to conduct a more comprehensive

evaluation of the multiple influences on reliability (Brennan, 2000; Cronbach et al., 1972; Hill, Charlambous, & Kraft, 2012; Shavelson & Web, 1991). In a generalizability study, or a G-study, the researcher uses analysis of variance to decompose score variance into its true score variance and the various components of the error variance. Ideally, the person component (test taker) ought to have the greatest percentage of variance (the true variance). This tells the researcher that the score reflects variance in test takers versus variance in contextual factors such as the rater, the task selection or the occasion of testing. By partitioning the error components, the researcher can better understand the testing conditions that impact the reliability of the scores. This allows the researcher to conduct decision studies, or D-studies, in which the researcher maximizes conditions that lead to greater reliability in scores. Although the utility of G-theory is well established in the measurement community, it has not been widely used in the context of teacher performance assessments. Hill et al. (2012) issued a call for increased use of G-theory in this context and reported results of a study which will be discussed in the following section of this literature review. The remaining portion of this current section will be devoted to what can be learned about performance assessment reliability through studies that have been conducted in contexts outside of teacher preparation. This review is not meant to be exhaustive, but rather to provide an overview of performance assessment conditions that impact reliability.

Gulikers, Baartman, and Biemans (2010) examined the implementation of competence-based assessments in two agricultural vocational training institutions. Using qualitative analysis of group interviews, they found that explicitly addressing assessment characteristics with all stakeholders (university teachers, students, and employers) was

valuable. In the absence of specially trained raters and a knowledgeable facilitator, the stakeholder ratings were unreliable. Overall, this study demonstrated the importance of considering the larger testing environment. The extent to which the goals and purposes of the assessment are transparent to the various stakeholders in the testing process impacts the extent to which reliable scores can come from that system.

The specific impact of rater on score reliability was explored in a meta-analysis by Hoyt and Kerns (1999). Selecting studies conducted in the context of psychotherapy and job evaluation, they found that about 37% of score variance could be attributed to rater bias. This finding warrants attention because it indicates that raters often introduce significant sources of noise in performance assessment scores. Clearly, adequate training of raters is one way to reduce this bias, but rater reliability is also influenced by the cognitive complexity of the scoring tasks. In a literature review on the use of rubrics in performance assessments, Jonsson and Svingby (2007) found that analytic, topic-specific rubrics that contained exemplars enhanced the reliability of scores.

In addition to the influence that raters and rubrics have upon score reliability, there is a considerable body of research that documents the effect that task selection has upon reliability (Brennan, 2000). Shavelson, Baxter, and Pine (1991) in an oft-cited study demonstrated that both task and methods heterogeneity severely limited the reliability of science performance assessments. In the context of a simulation-based acute care skills assessment for medical students, Boulet (2003) found that the specificity of the case (the particular acute care scenario) accounted for the greatest portion of score variance—greater than the variance due to person (test taker). It took six cases and four raters to reach a reliability coefficient of .74. This study serves as a caution to those who

design performance assessments; unless a G-study is conducted one cannot be certain that any particular number of sampled tasks is sufficient.

In contrast to Boulet's (2003) finding, Conigliaro and Stratton (2010) found only 1% of the variance in their study was due to occasion (similar to case selection in the Boulet study). The Conigliaro and Stratton study was designed to evaluate the use of performance assessments of medical education faculty who were conducting medical rounds. These researchers also found that 64% of the variance was attributed to person, and 35% of the variance was found in the rater-nested-within person and occasion facet. Because the G-study had a partially nested design (each rater was assigned to some, but not all persons, for some, but not all occasions) it was not possible to parse fully the role of rater. It is surprising that occasion accounted for such a small amount of the variance; it is possible that the simple rubric (15 items, 3 dichotomous and 12 trichotomous) influenced the stability of the occasion facet. In addition, based on this simple rubric, it seems clear that the medical faculty were expected to engage in very similar behaviors from occasion to occasion. In contrast, in the Boulet study, the medical students were expected to engage in very different behaviors depending upon the particular acute care scenario. Developers of performance assessments for teaching candidates ought to reflect upon this pair of studies and consider the extent to which teachers must access different knowledge and skills in each teaching scenario. TPPs may need to sample a fairly large number of performance assessments in order to reach "acceptable" ( $\approx .80$ ) levels of reliability, not to mention the higher levels ( $\approx .90$ ) that are fitting for high-stakes decision-making.

The studies that have been presented in this section were selected because of their potential to shed light on issues pertaining to the reliability of performance assessments in general. Taken together, they suggest that transparency of purpose to all stakeholders, rater training, rubric design and task sampling are key factors that impact reliability and that G-theory is an efficient means to determine the extent to which each facet influences the reliability of any given study.

This background information on reliability, together with information from the previous sections on the ways in which performance assessments are defined and in which their scores are validated provide an informed lens through which one can evaluate the following empirical studies related to the use of performance assessments in TPPs.

**Performance assessments in the context of TPPs.** In this section, I discuss studies that addressed the use of performance assessments in the context of teacher preparation programs. I have organized this evidence around the following themes: initial adoption of the performance assessment, issues related to validity, and issues related to reliability.

***Initial adoption of a performance assessment.*** In this section, I provide analysis on three studies that took place in the context of the initial adoption of a performance assessment. Peck, Galluci and Sloan (2010) used case study methodology to analyze how one TPP moved from resisting to embracing the high-stakes performance assessment mandated by the state of California. Study participants ( $n= 35$ ) were faculty and staff of a 13-month initial licensure master's degree program. Over a period of 18 months, researchers collected qualitative data in the form of field notes, semistructured interviews, freewrites, and artifacts such as course syllabi and program assignments.

Despite faculty members' initial resistance to and distrust of the mandated performance assessment of their students, their decision to use the new mandate as an opportunity for inquiry resulted in programmatic changes that were viewed favorably by faculty members. Although this study provides an example of how teacher educators can embrace the use of mandated performance assessments, the quality of the performance assessment itself was not critically evaluated. As such, this qualitative study says more about the positive nature of a faculty that takes an inquiry stance toward program development than about the consequential basis of the validity of a mandated, high-stakes performance assessment.

Wentworth, Erickson, Lawrence, Popham, and Korth (2009) described one TPP's efforts to create a Clinical Practice Assessment System (CPAS) in response to accreditation requirements. Although this study attended to some of the measurement issues involved in using a new assessment (in contrast to the previous study), the effort was weak, at best. The researchers created 10 subscales with 44 total indicators based upon their understanding of the Interstate New Teacher Assessment and Support Consortium (INTASC) standards. Although they computed Cronbach's alphas as measures of internal reliability for the 10 subscales, they did not use factor analysis to validate the subscale structure of the instrument. They only had one rater score each performance and reported little opportunity to train these raters. Because they did not use generalizability theory to evaluate the reliability of their scores, and because other studies have shown significant issues with rater bias (Hoyt & Kerns, 1999), the reliability of their scores is suspect. Wentworth et al. provided anecdotal, descriptive evidence that CPAS "provides teacher candidates and supervisors with greater consistency in expression of



expectations, a common language for communication, and a basis for evaluation, as well as a history of teaching performance” (p. 20). This anecdotal information may indicate fruitful areas of future research, but the researchers have made claims that their data do not support. The next study that I present could serve as a model of how the Wentworth et al. study might be improved.

Also in response to California’s mandate for performance assessments in TPPs, Riggs, Verdi and Arlin (2009) conducted a local analysis of the validity and reliability of the state-approved version of the California Teachers Performance Assessment (TPA). In the state-approved form of the TPA, there are four tasks; each task is evaluated with a global 4-point rubric—even though there are multiple subindicators for each task. Furthermore, a modal score, rather than a mean score, is used to make pass/fail decisions on the assessment. When local teacher educators conducted their own validity and reliability studies, they found that the use of global and modal scores limited the information that teacher candidates could potentially receive from the TPA. Through in-depth analysis of inter-rater reliabilities and intra-class correlations, they found significant evidence of lack of reliability of scores. In response, the researchers put new reliability safeguards in place: initial rater training, continual monitoring of scores, and periodic rater evaluations. A potential weakness in this study is the fact that data were collected from volunteers during a pilot study. Also, a generalizability study would have provided even more extensive information on reliability. Nevertheless, this work is a model to other TPPs as it is in line with the understanding that validity and reliability analyses ought to be an ongoing and cumulative effort (AERA, APA, & NCME, 1999).

The studies presented in this section provide examples of three different stances that can be taken in the initial adoption of high-stakes TPP performance assessments required for licensure or accreditation. The first stance is a relational and dispositional stance that evaluates the perceptions of the new performance assessment (Peck et al., 2010). The second stance is one in which superficial attention is paid to the most basic measurement qualities of the TPP (Wentworth et al., 2009). The third stance is one in which the TPP takes full ownership of the valid and reliable use of the performance assessment scores (Riggs et al., 2009). In the next section, I present studies that specifically address aspects of performance assessment validity.

***Validity in TPP performance assessments.*** Keeping in mind that validity is a unitary construct (Messick, 1981), the following three articles provide information, albeit limited, on aspects of validity that ought to be considered in an implementation of a TPP performance assessment. Messick (1989) argued for the importance of attending to the consequential basis for test score validity, and the first two studies attended to this basis. Bunch, Aguirre, and Tellez (2009) conducted a qualitative analysis of the opportunity that the Performance Assessment of California Teachers (PACT) provided for teacher candidates ( $N=8$ ) to reflect on their knowledge of effective mathematics teaching strategies for students who are English language learners. The authors neglected to communicate sufficient detail on their qualitative methodology, and so it is difficult to evaluate the rigor of their study. However, they presented evidence that their participants engaged in in-depth reflection on how to best meet the needs of their English language learning students. The authors remarked “it is hard to imagine traditional paper and pencil assessment promoting these kinds of deep and broad discussions about preservice

teacher candidates' developing knowledge and skills related to the instruction of [English language learners]" (pp. 122-123).

Whereas Bunch et al. (2009) looked at the consequential basis of validity from the perspective of teacher educators, Okhremtchouk, Seiki, Gilliland, Ateh, Wallace and Kato (2009) looked at the intended and unintended consequences of the performance assessment from the perspective of the teacher candidates. Okhremtchouk et al. used sequential exploratory design as a qualitative research model and grounded theory as the method of analysis to look at open-ended surveys of 73 teacher candidates who completed the PACT during the 2006-2007 year. They found that 44% of candidates rated the PACT as being helpful to their student teaching practices. Sixty-five percent of respondents said that the PACT negatively impacted their university coursework, and 94% said that the PACT negatively affected their personal time and life. In addition, 66% indicated that the specific implementation context of the PACT was unhelpful.

The juxtaposition of the Bunch et al. (2009) and Okhremtchouk et al. (2009) reveals the necessity of examining the consequential basis of test validity from the perspectives of multiple stakeholders. The very same assessment may have negative consequences from the perspective of one stakeholder and positive consequences from the perspective of another stakeholder.

Sandholz and Shea (2012) examined whether university supervisors could predict preservice teacher scores on the PACT. Prior to making the predictions, each university supervisor had completed three cycles of classroom observations that included detailed, formative feedback. The correlation between predictions and total scores was .289. The authors had hypothesized that predictions might be better for extremely high or low

candidates, but this, in fact, was not the case. The predictions were no more accurate for the extreme scores. Of the supervisor predictions that did not match the PACT scores, approximately half under-predicted and half over-predicted. The lack of agreement between supervisor and PACT scores highlights the difficulty of designing effective and convincing validation studies.

To summarize this section, all three articles presented here serve to highlight the challenge of designing studies that can contribute to coherent arguments for validity of performance assessment scores. The first two articles highlighted the importance of the consequential aspect of validity, and the last article highlighted the need for careful consideration of what can be deemed to be evidence of concurrent or construct validity. Moving from validity to reliability, the next section of this review will synthesize work that has been done on the reliability of scores from performance assessments in TPPs.

***Reliability in TPP performance assessments.*** Hill, Charalambos, and Kraft (2012) provided an eloquent and convincing argument for attending to observational *systems* within the context of a performance assessment. Instead of having a view of reliability limited to inter-rater reliability, Hill et al. argued for attention to be paid to the instrument itself, to the raters, the rater training, the number of observations, etc. Furthermore, they recommended generalizability theory as an effective means of evaluation of the entire system of observation.

Hill et al. (2012) also presented findings of a study in which they examined the reliability of scores from the Mathematical Quality of Instruction (MQI) observational assessment. With respect to the reliability of their scores, the results were ambiguous at best. The teacher facet, the source of true score variance, accounted for between 27%

and 46% of the score variance. Hill and colleagues conducted several decision studies and found that it would take 3 occasions and two raters to reach reliability scores of .77, .71, and .81 on the MQI subscales. It is also important to note that Hill et al. chose to compute relative reliability coefficients instead of absolute, or criterion based, coefficients. This means that the reliability values pertain to ordering a set of scores, not to making criterion-based decisions about scores. Overall, this study illustrates how expensive, in terms of times and resources, that it can be to create an assessment system with adequate reliability of scores.

Paetorius, Lenske, and Helmke (in press) also used generalizability theory to examine the role of observer ratings of instructional quality. Twelve trained raters scored 57 instructional sequences that were approximately 10 minutes in length, 390 untrained raters viewed 3 instructional sequences. Depending on the particular measure examined, 16-44% of score variance was due to instructional quality, while 12-40% was due to rater bias. Although the trained raters experienced an eight-hour training, their bias was comparable to that of the untrained raters. This is a surprising finding. However, the study also included a qualitative portion in which researchers prompted raters to provide explanations for their scores. Researchers reported that the trained raters were able to provide significantly more manual-based explanations for their scores. Thus researchers interpreted the trained raters' scores to be more valid than the scores of the untrained raters. Nevertheless, it is a puzzling finding that untrained raters had the same or less bias than trained raters. One confounding factor could be that the untrained raters watched fewer videos in a fixed sequence. Additional studies are necessary to parse out the effect of video sequence. Another potential weakness of the study may be the instrument itself.

The instrument measured two aspects of instructional quality, classroom management and personal learning support, irrespective of content. It could be that the items to be scored required high levels of inference and the training was insufficient to prepare raters for this level of inference. Subject-specific instruments may have advantages here. In addition to providing an example of an instance when training did not help to reduce rater bias, this study provides another example of how a G-study can be designed.

Pianta and Hamre (2009) presented a comprehensive argument for the need for standardized observation protocols. They argued that scores from such observation have the potential to be shown to be reliable through G-theory, to be validated for predicting student learning, and to be effective in promoting positive change in teaching. Although they did not provide a full report of their empirical work, they did report that rater effects in their study were between 4 and 14% of score variance. They also reported evidence that the quality of classroom instruction decreased over the course of the school day. This is an important source of error that has not been widely attended to in classroom-based observational assessments. To conclude this section on reliability of performance assessment scores in the context of teacher preparation programs, the three studies reported here each demonstrate the power of G-studies to provide information about the reliability of scores from performance assessments.

### **Design-Based Research**

I close this chapter with a review of literature on the methodological approach that I have chosen to utilize in my study. My intent for this review is not to provide an exhaustive review of all articles published on design-based research (DBR), but to

articulate the major developments on the use of DBR in education. In particular, I focused this review on articles that addressed indicators of quality in DBR.

**Terminology and origins.** There have been several terms used to indicate design-based research. These include: *design research*, *design experiments*, *design experimentation*, *design-based methods*, *design studies*, and *development research*. Table 1 provides a list of constitutive definitions of DBR from the literature base.

In the early 1990's the work of Ann Brown and Allan Collins launched DBR as a visible and valuable method of conducting educational research. Certainly DBR had been utilized in education prior to their work, but these researchers clearly articulated the methodological approach and advocated for its use in educational research. During the following decade, the Design-Based Research Collective was formed, and there were three special issues of educational journals that focused on DBR: *Journal of the Learning Sciences*, vol. 13(1), *Educational Researcher*, vol. 32(1), and *Educational Psychologist*, vol.39(4).

**Elements of DBR.** Brown (1992) was instrumental in articulating design experiments for the field of educational research. Although this paper was published two decades ago, Brown was able to anticipate and describe the key advantages and challenges of DBR that continue to be discussed to this day. Brown identified a critical tension in the purpose of design research as one of balancing contributions to learning theory as well as contributions to classroom practice. Because design research is meant to inform practice, the educational interventions that are the subject of study must be able to be carried out in real classrooms with levels of support that are available to the typical teacher in the typical school.

Table 1

*Constitutive Definitions of Design-Based Research*

Source	Constitutive definition of Design Research
Anderson & Shattuck (2012)	DBR is a methodology designed by and for educators that seeks to increase the impact, transfer, and translation of education research into improved practice. In addition, it stresses the need for theory building and the development of design principles that guide, inform, and improve both practice and research in educational contexts. (p. 16)
Barab & Squire (2004)	Design-based research is not so much an approach as it is a series of approaches, with the intent of producing new theories, artifacts, and practices that account for and potentially impact learning and teaching in naturalistic settings. (p. 2)
Cobb, Confrey, diSessa, Lehrer, & Schauble (2003)	Prototypically, design experiments entail both “engineering” particular forms of learning and systematically studying those forms of learning within the context defined by the means of supporting them. This designed context is subject to test and revision, and successive iterations that result play a role similar to that of systematic variation in experiment (sic). (p. 9)
Collins, Joseph, & Bielacyc (2004)	Design experiments were developed as a way to carry out formative research to test and refine educational designs based on theoretical principles derived from prior research. (p. 15)
Confrey (2006)	A design study is an extended investigation of educational interactions provoked by use of a carefully sequenced and typically novel set of designed curricular tasks studying how some conceptual field, or set of proficiencies and interests, are learned through interactions among learners and with the guidance of an instructor or form of tutor. The study seeks to document what resources and prior knowledge the students bring to the task, how students and teachers interact, how records and inscriptions are created, how conceptions emerge and change, what resources are used, and how teaching is accomplished over the course of instruction, by studying student work, video records, and classroom assessments. (p. 2)
Design-Based Research Collaborative (2003)	Design-based research (Brown, 1992; Collins, 1992) is an emerging paradigm for the study of learning in context through the systematic design and study of instructional strategies and tools. We argue that design-based research can help create and extend knowledge about developing, enacting, and sustaining innovative learning environments. (p. 5)  We do not claim that there is a single design-based research method, but the overarching, explicit concern in design-based research for using methods that link processes of enactment to outcomes has power to generate knowledge that directly applies to educational practice. (p. 7)
Shavelson, Phillips, Towne, & Feuer (2003)	Such research, based strongly on prior research and theory and carried out in educational settings, seeks to trace the evolution of learning in complex, messy classrooms and schools, test and build theories of teaching and learning, and produce instructional tools that survive the challenges of everyday practice. (p. 25)



Brown (1992) rejected the assumption that laboratory studies, or highly controlled experimental studies, were readily transferrable to the complex ecological environment of the typical classroom. She believed that both the intervention and the entire ecological system needed to be studied simultaneously. In the following passage Brown explains her systemic approach and the careful attention to assessment that is necessitated by this approach:

Thus, we are responsible for simultaneous changes in the system, concerning the role of students and teachers, the type of curriculum, the place of technology, and so forth. These are all seen as inputs into the working whole. Similarly we are concerned with outputs from the system, a concern that leads us to look at new forms of assessment. It is essential that we assess the aspects that our learning environment was set up to foster, such as problem solving, critical thinking, and reflective learning. Assessment also allows us to be accountable for the results of our work to the children themselves, to parents, to teachers, to local authorities, and last but not least, to fellow scientists (p.143).

In this passage, Brown clearly ties the design of assessment to the success of a DBR study. It is the carefully crafted assessment that will be sensitive to the change that the DBR study produces.

In addition to Brown (1992), Collins (1999) and Collins, Joseph, and Bielzyc (2004) helped to define DBR by contrasting the characteristics of DBR studies with those of laboratory studies. Table 2 shows these contrasts.

Building on Brown's (1992) emphasis on the importance of assessment, Collins et al. (2004) suggested that there are at least three types of dependent variables that ought

to be assessed in a DBR study. These are: 1) climate variables such as engagement, cooperation, risk-taking, and student control; 2) learning variables such as content knowledge, dispositions, learning strategies, and metacognitive strategies; and 3) systematic variables such as sustainability, scalability, ease of use, and cost.

Table 2

*Contrasts Between Design Experiments and Laboratory Studies*

Laboratory Study	Design Experiment
<b>Laboratory Setting</b> Learners have no distractions, materials are well defined and present in a standardized manner.	<b>Messy Situations</b> Real life learning avoids distortions of laboratory experiments.
<b>Single Dependent Variable</b> Researcher is interested in one clearly defined variable.	<b>Multiple Dependent Variables</b> There are many variables that matter, but researcher may not be able to attend to all.
<b>Controlling Variables</b> The goal is to identify a few independent and dependent variables and to hold all other variables constant.	<b>Characterizing the Situation</b> The goal is to identify all variables that affect any dependent variable of interest.
<b>Fixed Procedure</b> Researchers follow a defined procedure that can be replicated by other researchers.	<b>Flexible Design Revision</b> Design researchers start with plans that are not completely defined and which are revised depending upon success of implementation.
<b>Social Isolation</b> In psychological learning experiments, researchers attempt to isolate a learner and minimize interaction with other learners.	<b>Social Interaction</b> Design experiments take place in complex social environments.
<b>Testing hypotheses</b> Researcher has one or more hypotheses that are systematically tested.	<b>Developing a Profile</b> The goal is to evaluate many different aspects of the design and develop a profile that characterizes the design in use.
<b>Experimenter</b> The experimenter makes all decisions about design and analysis of the experiment.	<b>Co-participant Design &amp; Analysis</b> Design researchers make an effort to involve different participants in order to include their expertise in producing and analyzing the design.

With regard to independent variables, Collins et al. (2004) suggested that the following list of independent variables be attended to: 1) setting, 2) nature of the learners, 3) required resources, 4) professional development, 5) financial requirements, and 6)

implementation path. They further explained that there is a web of interrelationships between these independent variables and the dependent variables. In trying to understand this web, DBR often involves the collection of voluminous amounts of data. Collins et al. make it clear that rarely is there enough time or resources to analyze all of the data that is collected.

**Challenges and indications of quality.** Brown (1992) identified two key threats to validity in DBR studies. Because DBR produces vast amounts of data, the researcher's selection of data to analyze becomes very important. Brown uses the term *Bartlett Effect* to indicate the risk of the researcher engaging in biased selection of the data. In a related vein, Shavelson, Phillips, Towne and Feuer (2003) provided a sharp critique of design-based studies that they believed had relied heavily on narrative accounts to make unwarranted claims. In an effort that appeared to be aimed at reigning in sloppy research, Shavelson et al. reminded design-based researchers to follow the guidelines set forth in their recently published *Scientific Research in Education* (NRC, 2002). Specifically they argued that design researchers, like all education researchers, should:

- pose significant questions that can be investigated empirically,
- link research to relevant theory,
- use methods that permit direct investigation of the questions,
- provide a coherent and explicit chain of reasoning,
- attempt to yield findings that replicate and generalize across studies, and
- disclose research data and methods to enable and encourage professional scrutiny and critique (p. 26).

The concern regarding biased selection of data on the part of the researcher is counterbalanced by the benefit of serendipitous findings that the researcher selects and that provide unanticipated insights. diSessa and Cobb (2004) provided an alternative perspective to Shavelson et al. (2003) in this passage:

It is manifestly impossible to study everything that happens in a design experiment, and trying to do so exhibits lack of scientific focus. Systematic analysis has its virtues, but researchers have finite time and hopefully the wisdom to study mostly issues that are well prepared and tractable. On the other hand, surprising things essentially always happen, and some data collection beyond core focus is always a good idea, pending trade-offs of time and effort with other tasks. In our experience, much of that data never gets analyzed. Many times we decide in advance on principles of data selection. (p. 87)

The scholarship cited here has informed DBR researchers of the benefit, dangers, and inevitability of researchers only analyzing part of the data collected in a DBR study. A balanced approach that is cautious to avoid biased selection and also open to serendipitous findings is one that must involve transparent communication of data collection and data analyses strategies. Brown (1992) indicated a willingness to store the full dataset from a DBR study in order to make it available to other researchers who would like to analyze the data in a different method (Shoenfeld, 1992).

In addition to the Bartlett Effect, Brown (1992) also acknowledged the risk in DBR studies of the Hawthorne Effect—that an intervention can be seen to have a positive effect simply because of the attention that the research team gives to the participants. However, Brown also pointed out that the Hawthorne Effect would produce improvement

indiscriminately, but the DBR researcher ought to be able to predict where the improvement will occur. This ability to predict improvement is a result of DBR studies being firmly grounded in theory.

## **Discussion**

The purpose of this section is to summarize and synthesize the research that has been reviewed in this chapter. I highlight key points from each of the five areas of review and discuss implications for this study.

**Preparing practitioners for EBP.** Because I have chosen to utilize social work researchers' process definition of EBP (Gibbs & Gambrill, 2002; Rubin, 2007, 2008; Rubin & Parrish, 2007) and five-step conceptual framework for EBP (Gibbs & Gambrill, 2002; Parish & Rubin, 2011; Rubin, 2008; Shlonsky & Gibbs, 2004; Thyer, 2006), I reviewed the existing literature on preparing social work practitioners to engage in the process of EBP. Unfortunately no evidence could be found that linked the EBP training of social workers to client outcomes. This may be the result of methodological difficulties (Evidence-Based Medicine Working Group, 1992; Hatala & Guyatt, 2002) or the result of the relative immaturity of this area of research. Only recently have studies systematically examined EBP training (e.g., Parrish & Rubin, 2011) and its integration with fieldwork practica (e.g., Mullen et al., 2007). It seems reasonable that clinical outcome studies could only be accomplished after EBP training has been systematically studied and established. Although it seems logical that training practitioners to engage in the process of EBP would result in their clients receiving better care, this hypothesis has yet to be empirically tested.

Training on EBP can also be evaluated by looking at educational outcomes. Although there is a larger body of research in this area than on client outcomes, the findings are limited by the lack of validation of measures of educational outcomes (Flores & Argimon, 2007; Shaneyfelt et al., 2006). Despite this weakness, there are some important patterns in the findings. First, clinical integration of EBP training has been more effective than didactic, classroom-based training (Coomarasamy & Khan, 2003; Khan & Coomarasamy, 2006). Training focused on specific EBP steps, such as asking a research question and finding relevant research, has been shown to increase skills in these areas (Cheng, 2003). The combination of didactic and hands-on training on all five steps has also been shown to increase the skills and knowledge of the EBP process (Parrish & Rubin, 2011).

Nevertheless, research has also pointed to the difficulty of maintaining the benefits of training over time (Cheng, 2003) and the challenge of changing the practitioner's post-training behavior (McCluskey & Lovarini, 2005). The field has yet to determine what types of follow-up measures are necessary so that social work practitioners maintain knowledge and skill gained from the initial EBP training.

In the university-based preparation of social workers, the field placement has been identified as key to EBP training and the advancement of EBP in social work practice (Edmond et al., 2006; Franklin, 2007; Mullen et al., 2007; Proctor, 2007; Rubin, 2007). Field instructors have endorsed positive feelings toward EBP, but do not regularly engage in EBP in their own practices (Edmond et al., 2006). In order to address the need to integrate EBP training into field work placements, researchers recommended EBP training for field supervisors (Edmond et al., 2006), training agency teams to engage in

EBP, and incorporating agency training into regularly scheduled trainings (Mullen et al., 2007).

There are several implications from this literature review for EBP training in the context of teacher preparation. First, as in all areas of research, there is the need for validated instruments. In the case of EBP training, the Evidence-based Practice Process Assessment Scale (EBPAS) (Parrish & Rubin, 2011) is a potential model of an instrument that measures EBP skills, attitudes, and behaviors and that has produced scores shown to be valid and reliable. Another implication is that hands-on, clinically integrated training is superior to didactic training alone. As in social work, the fieldwork placement in education may be key in the development of novice teachers' ability to engage in the process of EBP.

**Linking research to practice in teacher preparation.** After gaining an understanding of EBP training in the field of social work, I reviewed the literature on linking research to practice in teaching and teacher preparation. No research could be located that explicitly defined or examined the process of evidence-based practice in the context of classroom teaching. Furthermore, education researchers have not clearly differentiated between the two interpretations of the phrase EBP: (1) EBP as an indication that a teaching practice has been supported by empirical research or (2) EBP as a *process* in which teachers engage to identify the best teaching practices to meet students' needs.

The first EBP definition (a practice supported by research) was used implicitly in many of the reviewed studies (Antil et al., 2003; Fuchs & Fuchs, 2001; Klinger et al., 2003; Lamb et al., 2007; Miller et al., 2005). These studies documented some poor

outcomes. Teachers overestimated their ability to implement a teaching strategy supported by research (Antil et al., 1998). Despite extensive training and follow-up support on research-based teaching strategies, implementation rates and fidelity rates were low (Klinger et al., 2003). In addition, teachers used research to validate their own teaching strategies and biases instead of reconsidering their practice in light of the research (Civil, 2011; Kennedy, 1997). The poor outcomes reported here may be artifacts of the limited view of EBP as a mere intervention instead of a dynamic process initiated by the teacher.

There were three studies in the review that, although not explicit in language, engaged teachers in the process of EBP. Sinnema et al. (2011) described collaborative inquiry between researchers ( $n=6$ ) and classroom teachers ( $n=26$ ) in which teachers were successful in finding and applying research relevant to classroom teaching concerns. The findings appeared promising but their use is limited because the researchers reported very little detail on the instruction that the teachers received and only reported details on the improved instruction of 2 out of the 26 cases.

Harrison et al. (2006) reported on a research methods course for early childhood teachers. Participants ( $N=70$ ) reported that reading education research was of little use to them without its practical application in their research projects. This finding is in line with findings from social work that clinically-integrated EBP training is superior to didactic training alone. The instrument used to measure students' perceptions about research and research use was not validated and no information about score reliability was provided. This limits the interpretations of the findings.



Everett et al. (2008) integrated science content instruction with an action research project in which preservice teachers ( $n=144$  groups of teachers over 9 semesters) engaged in finding and applying research to classroom teaching. The fact that 96% of the groups were successful at linking research to their teaching practices was very promising. The lack of detail on the instruction and support provided to the preservice teachers limits the replicability of this study.

The three previous study citations serve as proof-of-concept articles; teachers and preservice teachers are capable of engaging in the process of linking research to practice. However, none of the studies explicitly discussed the five-step EBP process and none provided the detail on intervention or instrumentation that would be needed to replicate the study findings. This points to the need for systematic and transparent research on this topic.

**Routines of practice.** As the previous section indicates, little research has been conducted on the process of EBP in the context of teacher preparation, yet the study of the process of EBP fits into a larger body of research on key practices of teaching that can be introduced in teacher preparation. Within the last decade, a new pedagogy of professional preparation (Grossman et al., 2009) has emerged within the field of mathematics education. Currently researchers are attempting to define routines of practice, also called *high leverage practices* (Ball et al., 2009), and to incorporate these practices into teacher preparation programs. Ball et al. identified eight criteria for a practice to be identified as a high-leverage practice. Table 3 lists these criteria and the ways in which the process of EBP fulfills the criteria.

Table 3.

*Ways in Which the Process of EBP Fulfills the High-Leverage Practice Criteria*

High-Leverage Practice Criteria	The Process of EBP
Supports work that is central to mathematics.	EBP provides a systematic way that teachers can learn the best methods for teaching math content.
Helps to improve the learning and achievement of all students.	Step 5 requires teachers to gather evidence to evaluate the effectiveness of their teaching. This would most often involve evidence of student learning.
Is done frequently when teaching mathematics.	Ideally, EBP would be conducted on an on-going and continual basis.
Applies across different approaches to teaching mathematics.	Because EBP depends upon teacher expertise, it is compatible with a teacher's unique approach to teaching mathematics.
Can be articulated and taught.	The five-step framework articulates the process and can be taught.
Is accessible to learners of teaching.	Existing research confirms that the steps of EBP are accessible to learners of teaching.
Can be revisited in increasingly sophisticated and integrated acts of teaching.	The five-step framework can be practiced in steps, outside of the classroom and gradually integrated into classroom teaching.
Is able to be practiced by beginners in their field-based settings.	Ideally, EBP is learned in a field-based setting.

The reviewed literature on routines of practice has the following implications for this study. In the training of preservice teachers, approximations of practice have been necessary in order to scaffold preservice teachers into the more complex behaviors of classroom teaching (Grossman et al., 2009). Similarly, the construct of *conceptions of quality* has been suggested to be useful in training preservice teachers to improve their performance in routines of practice (Moss, 2011). Finally, when teaching a routine of

practice, it has been suggested that teacher educators take responsibility to name the elements of the practice/activity, provide novices with the opportunity to practice the elements of the practice, and to assess the learning around these key elements (Kazemi et al., 2009).

These three responsibilities of the teacher educator have implications for the design of the current study. The teacher educator responsible for preparing preservice teachers to engage in EBP must define the five-steps in the context of classroom teaching, must provide opportunities to practice the steps and must be able to assess these practices. Because the assessment of a practice is, by definition, a performance assessment, this chapter included a review on this area of literature as well.

**Performance assessments in teacher preparation.** The purpose of this section of the literature review was to identify factors that contribute to valid and reliable scores from a preservice teacher performance assessment designed for this study. The review identified general factors related to validity and reliability of performance assessment and factors specific to performance assessments in teacher preparation programs. In general, there is a key distinction that must be made between performance and competence (Jonsson & Mattsson, 2011; Messick, 1984). Performances are context dependent and failure in a performance should not necessarily be interpreted as lack of competence. Research has shown that reliability is impacted by rater, task, occasion and rubric, among other factors (Boulet, 2003; Brennan, 2000; Conigliaro & Stratton, 2010; Gulikers et al., 2010; Hoyt & Kerns, 1999; Jonsson & Svingby, 2007; Shavelson et al., 1991). To identify and minimize sources of error, generalizability theory is a powerful analytic tool

and has been recommended specifically for the context of teacher performance assessments (Hill et al., 2012).

Specific to performance assessments in teacher preparation programs, research has shown that validation studies must be carefully considered. The perspectives of multiple stakeholders ought to be considered and may not align with each other (Bunck et al., 2009; Okhremtchouk et al., 2009). In addition, concurrent validation studies must be carefully designed as not all teaching performances produce similar data (Sandholz & Shea, 2012).

Reliability of teacher performance assessments is best studied in the context of *systems* of assessment and with the use of generalizability theory (Hill et al. 2012).

Rater bias has been shown to be a significant source of error in teacher performance assessments (Hill et al., 2012; Paetoriuos et al., in press) and may possibly be minimized through standardized observation protocols (Pianta & Hamre, 2009).

**Design-based research.** Because DBR was identified as the appropriate research design to develop the EBP intervention and system of assessment for teaching preservice teachers, it was important to review literature on the indicators of quality for this methodological approach.

A critical tension in the purpose of DBR is to balance contributions to learning theory as well as contributions to classroom practice (Brown, 1992). In this study, the tension is between defining the five-step process for the context of education and creating an effective intervention that can be used by teacher educators. The development of systems of assessment is key to balancing the tension in purposes and to determining the effectiveness of the intervention in a DBR study (Brown, 1992).

Unlike experimental studies that can be designed to isolate and find causal relationships between independent and dependent variables, DBR studies are designed to report on a web of interconnected independent and dependent variables (Collins et al., 2004). A quality DBR study reports on climate, learning, and systematic variables in order to articulate the context in which the learning outcome was achieved (Collins et al., 2004).

Brown (1992) identified two key threats to the validity of DBR studies: (1) the Bartlett effect in which the researcher engages in biased selection of data and (2) the Hawthorne effect in which improvement in outcome is due only to the attention of the researcher and not the intervention. Shavelson et al. (2004) provided several recommendations to mitigate the Bartlett and Hawthorne effects. Among the recommendations that will be used in this study are: (a) posing significant research questions, (b) linking research to theory, (c) using methods that permit direct investigation of questions, and (d) disclosing research data and methods. Adherence to these recommendations will contribute to the quality of this DBR study.

In conclusion, the information gleaned from this literature review has been essential to the design of the study described in the next section. The review on EBP training in social work provided guidance as to what has worked with training on the five-step EBP process for social work practitioners. The success of social work's clinically-integrated training has directly informed the intervention to be used in this study. The review on linking research to practice in teaching and teacher preparation provided context-specific information about the educational environment that also guided the development of the study. In addition to finding proof-of-concept studies that

demonstrated that preservice teachers are capable of linking research to practice, a significant gap was identified in the literature. To date, no studies have systematically studied the process of training preservice teachers to engage in the process of evidence-based practice. The literature on routines of practice in mathematics education helped to set this study into the larger context of research on preparing preservice mathematics teachers to engage in activities that will become central to their future teaching practices. This also highlighted the key responsibilities of teacher educators who aim to create interventions focused on routines of practice. The literature on performance assessments in teacher preparation provided insights about the development of an assessment that produces valid and reliable scores. Specifically, designing a system of assessment and using generalizability theory to assess for reliability are two key recommendations from this literature that were heeded in this study. Finally, the literature on indicators of quality for DBR provided guidance on how to maximize the success of this methodological approach. In the next chapter, I provide a detailed description of the study design and methodology.

### Chapter 3: Method

The overall purpose of this study was to identify how to prepare preservice elementary education teachers to engage in the process of applying research to their mathematics teaching. In Chapter 1, I developed the argument for the need for this study by looking at the problems of lack of mathematics achievement as well as disparity in mathematics achievement. These problems can be addressed from three different levels. At the student level, one asks, “What can students do to achieve more in mathematics?” At the teacher level, one asks, “What can teachers do to teach mathematics more effectively for every student?” At the teacher preparation level, one asks, “What can be done in teacher preparation programs to better prepare teachers to teach mathematics effectively for each and every student?” It is at this level of teacher preparation that I have focused my efforts.

Currently, the US government has mandated through legislation that teachers use evidence from research to guide their teaching decisions (e.g. NCLB, 2002). Yet, the *process* of applying research to practice in teaching has remained, largely, unstudied. Teachers have not consistently changed their practices due to the input of education research (OECD, 2009). Furthermore, mathematics teacher educators do not have a shared research base to inform their work to prepare teachers of mathematics in general (Cochran-Smith & Zeichner, 2005; Hiebert & Morris, 2009) let alone for the specific process of applying research to practice.

I assert that the meta-*process* of applying research to practice must be addressed in order for the gap between research and practice to begin to close. I have looked to the field of social work's five-step conceptual framework of the process of evidence-based practice (EBP) as a potential model for teaching teachers the process of applying research to practice. As was discussed in Chapter 1, social work researchers have defined EBP as the *process* of finding and evaluating research in the context of making practice decisions (Rubin, 2008). Based upon this *process* definition of EBP, social work researchers (e.g., Mullen, 2004, 2006; Rubin, 2008; Shlonsky & Gibbs, 2004; Thyer, 2006) have also developed a conceptual framework that explicates the steps involved in the EBP process (See Table 4).

Table 4

*Conceptual Framework for the Process of Evidence Based Practice*

---

Step 1	Formulate an answerable practice question.
Step 2	Search for the best research evidence.
Step 3	Critically appraise the research evidence.
Step 4	Select the best intervention after integrating the research evidence with client characteristics, preferences, and values.
Step 5	Evaluate the outcome of this practice decision.

---

This framework has been used successfully in the training of preservice and inservice social workers (Howard, Allen-Meares, & Ruffolo, 2007; Parrish & Rubin, 2011), and it was the conceptual foundation for the design of the intervention in this study. This study used the five-step conceptual framework as the basis for preparing preservice teachers to engage in the routine of EBP.



In Chapter 2, I reviewed literature in the five areas that are foundational to the design of this study. The review of literature on preparing social work practitioners to engage in EBP provided insight into the five-step framework and the advantage of clinically integrated training. The review of literature addressing the application of research to practice in the context of teacher preparation programs provided an empirical basis for understanding the unique context of teacher preparation as the site for work on linking research to practice. The review on high-leverage routines of practice helped to place this study in the context of a larger movement to reconceptualize teacher preparation in terms of practices that teachers will encounter in the classroom. The review of performance assessments provided foundational knowledge on the development of performance assessments and the use of generalizability theory to understand the reliability of performance assessment scores. Finally, the review of design-based research was critical to building the case for why this methodology is appropriate given the purpose of the study

In the present chapter, I provide a detailed description of the study design. The design is in direct response to the purpose of the study (Chapter 1) and the work that has been done previously in and around this topic of applying research to practice (Chapter 2). The overarching purpose for this study is to describe how mathematics teacher educators can prepare preservice elementary teachers to engage in the process of evidence-based practice. The specific objectives of this study are: (1) to use the conceptual, five-step framework for the process of EBP to create a teaching intervention that mathematics teacher educators can use to prepare preservice teachers to apply

education research to teaching practice; and (2) to create a system of assessment that supports the teaching intervention.

I decided to use design-based research (DBR) methods to address the overarching research objectives. DBR involves studying, simultaneously, the intervention and the ecological environment that supports the intervention. Unlike true experiments where the goal is to isolate the effect of the independent variable on the dependent variable, in DBR the goal is to understand a web of interconnected independent and dependent variables (Collins et al., 2004). DBR involves the collection of data that are both quantitative and qualitative in nature (diSessa & Cobb, 2004). These data speak both to the intervention and the environment that supports the intervention (Brown, 1992).

This study involved conducting the fifth iteration of the overall DBR study. To lay the foundations for the fifth iteration of the study, I provide a succinct discussion of the designs and findings from the first four iterations in order to make a case for the approach used in the fifth iteration. Then I describe the design of the fifth iteration in detail.

### **The First Four DBR Iterations**

Figure 4 provides an overview of the iterations that are involved in this DBR study. In this section, I discuss each of the first four iterations. This discussion includes descriptions of the interventions, the methods used to study the interventions, and the results of the analyses that were conducted. When appropriate, there will be an explanation of how the results from one iteration influenced revisions to the next iteration.

	Semester	Course (#of sections) Number of participants	Intervention	Evaluation
Iteration 1	Spring 2011	Math Methods I (3) n = 73	-ERP -One section experiences additional support from education librarian	1. Quasi-experimental study: Comparison of 2 levels of intervention with control group 2. Grounded theory analysis of ERP application essays 3. Treatment Potency Test
Iteration 2	Summer 2011	Math Methods I (1) n = 30	-ERP -Micro-Teaching x1 -5 additional research articles	1. Comparison on Information Literacy Skills Assessment 2. Treatment Potency Test
Iteration 3	Fall 2011	Math Methods II (1) n = 25	-ERP -Support from Librarian -Micro-teaching X2 and reflections -5 additional research articles	1. Comparison on Information Literacy Skills Assessment 2. Treatment Potency Test 3. Semi-structured Interviews (n=3)
Iteration 4	Summer 2012	Math Methods I (1) n = 30	-Revised ERP with rubric -Support from Librarian -Micro-teaching X2 -Pilot formative assessment	1. Comparison on Information Literacy Skills Assessment 2. Treatment Potency Test 3. Performance Assessment Pilot 4. ERP Full Rubric Pilot
Iteration 5	Fall 2012	Math Methods I (1) n = 12	-Revised ERP with rubric -Classroom portfolio -Pre/Post performance assessments	1. Comparison on Information Literacy Skills Assessment 2. Treatment Potency Test 3. Teaching Portfolio 4. Performance Assessment Generalizability Study

Figure 4. Overview of the iterations of the DBR agenda.

**Iteration 1.** I provide detailed information about Iteration 1 because it provided the foundation for subsequent iterations. In particular, I provide substantially more detail on the design of this iteration than on Iterations 2 through 4. My report on Iteration 1 consists of the following sections: design of the intervention, description of the ecological environment, and avenues of inquiry.

*Design of the intervention.* In the first iteration of the DBR agenda, I used the five-step EBP framework (see Table 4) to design a teaching intervention to engage preservice elementary teachers of mathematics in a scaffolded journey of applying research to their mathematics teaching. I titled this intervention the Education Research Project (ERP) (see Appendix A for the original formatting of the project). This project required preservice teachers to engage in seven tasks that approximated the five-step EBP process. Table 5 shows the correspondence between the seven ERP tasks and the five steps of the process of EBP.

The ERP was designed to introduce preservice teachers to a process of applying research to teaching and to give them opportunity to practice an approximation of that practice. The design of this project reflected the fact that I defined the process of EBP as a high-leverage routine of practice (Ball et al., 2009) for mathematics teaching, and that I engaged in a pedagogy of practice (Kazemi et al., 2009) in my instruction of preservice teachers.

Based upon previous experience working with preservice teachers, I anticipated that the ERP would be challenging for preservice teachers, and I wanted to explore two levels of this intervention. For Level One of the intervention, preservice teachers completed the ERP as described above. For Level Two of the intervention, the

Table 5

*Correspondence Between Tasks in the ERP and the Five-Step EBP Framework*

ERP Task	EBP Step
1) Identify a mathematics classroom-based problem or question	1
2) Turn this question into a researchable question	1
3) Identify search terms and education databases	2
4) Conduct a search of the literature and find an empirical, peer-reviewed article that addresses the research question	2
5) Read the article	2
6) Write a summary of the article	3
7) Write an essay about how the information from the article could be implemented in the preservice teachers' future teaching.	4, 5

university's education research librarian provided additional support for ERP activities one through four. The librarian co-taught two sessions. The first 75-minute session focused on how to turn a classroom based problem into a question that can be answered through a search of research literature. The second 75-minute session was conducted in a workshop format in a library computer lab and was designed to assist preservice teachers in their search for research articles.

***Ecological environment.*** IRB approval was obtained to engage in this study.

The study took place at a large, public metropolitan university in the southeastern United States. Participants were preservice elementary education teachers who were completing their mathematics education course requirements. Participants were recruited from three sections (A, B, and C) of Mathematics Methods 1, the first of a two-semester sequence of mathematics methods courses for elementary school teachers. All three sections of the

Mathematics Methods 1 course were taught by doctoral level graduate students. The Mathematics Methods I course curriculum provided an introduction to teaching elementary level mathematics and included emphasis on problem solving, whole number concepts, and rational number concepts.

Students enrolled in sections A and B were required to complete the ERP as a part of their normal course requirements. By volunteering to participate in the study, preservice teachers consented to have their work analyzed for research purposes in addition to the normal analysis as part of the course. Enrollment was voluntary and had no impact on course grading. Potential participants were informed of the study and given at least two days to decide if they wanted to participate before being asked to sign consent forms. Prior to enrolling in section A, B, or C, none of the students knew if or which section would be completing the ERP because it was a new assignment.

*Avenues of inquiry.* I engaged in three avenues of inquiry during Iteration 1: a quasi-experimental study that compared the two levels of intervention to a control group, a grounded theory analysis of the ERP essays on application of research to practice, and a treatment potency test.

*Quasi-experimental study.* Consistent with the goals of DBR studies, I wanted to understand the conditions that facilitate preservice teacher success with the ERP. In the quasi-experimental study, I compared outcomes from three sections of the mathematics methods course. Section A experienced Level Two of the intervention—librarian assistance with the ERP. Section B experienced Level One of the intervention—no librarian assistance with the ERP. Section C served as the control group, and students in this section did not complete the ERP.

Two tools were piloted during Iteration 1. The first was the Information Literacy Questionnaire (Appendix B) that gathered data on: 1) participant demographics and history related to teaching and information literacy, 2) intentions to use research to inform future teaching practice, and 3) information literacy skills. This questionnaire was given to each of the three course sections at the beginning and end of the semester. Because the underlying structure of the questionnaire had not been validated, total score was not calculated or analyzed. Instead, change in individual item scores from pretest to posttest was evaluated.

The second tool was a rubric (Appendix C) for the evaluation of the first four tasks of the ERP from Sections A and B. This rubric was designed to evaluate proficiency in Standard One of the Association of College and Research Libraries (2000).

Overall, the pre-post change scores on individual questionnaire items indicated that there was some benefit to the completion of the education research project. For example, a repeated measures ANOVA was run on the scores from question 8D that asked students to rate the likelihood of using scholarly articles to inform future teaching (see Table 6 for descriptive statistics on this item). Although the ANOVA showed no differences in gains between group scores  $F(2, 72) = 2.69, p > .05$ , a *t*-test statistic was computed for A and B (treatment) being contrasted with C (control). The obtained *t* value was 2.25,  $p < .05$ , and so there was evidence that the groups that completed the research project (A and B) reported greater gains in perceived likelihood of using scholarly research articles to inform their teaching practices.

In addition to comparing groups A and B to the control group, I also conducted analyses to compare group A (level 2 of the intervention) with group B (level 1 of the

intervention). A full report of this comparison can be found in van Ingen and Ariew (manuscript in preparation), and a succinct summary is described below. The first four tasks of the ERP were scored independently by two researchers with the rubric in Appendix C.

Table 6.

*Results Reported by Section from Information Literacy Questionnaire Item 8D*

When you think about your future teaching career, how likely do you think you are to use scholarly research to help inform your teaching? Rank on a scale of 1 to 5. (1=Not at all likely, 3= Somewhat Likely, 5=Very Likely)						
Section	Pretest		Posttest		Pre-Post Difference	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
A ( <i>n</i> =31)	3.90	0.94	3.97	0.87	0.07	0.96
B ( <i>n</i> =24)	3.67	0.96	4.08	0.78	0.42	0.97
C ( <i>n</i> =18)	3.71	0.92	3.41	0.71	-0.29	0.77

Based on this initial scoring, the rubric was revised, and again the two researchers independently scored the ERP tasks with the revised rubric (Table 7). As indicated on this rubric, the items were combined to form three indices: Search Question Index, Search Index, and Article Index.

Table 8 provides the descriptive statistics for the three indices for groups A (*n*=28) and B (*n*=24). Prior to conducting inferential multivariate statistics, the data were screened for violations to assumptions of a one-way multivariate analysis of variance (MANOVA). Due to study design, there was no indication of lack of independence of observations. There was limited evidence of lack of univariate normality (see Table 8 skewness and kurtosis values). However, these were relatively minor violations and were not deemed sufficient enough to warrant an approach other than a MANOVA, which is relatively robust to minor violations of normality (Stevens, 2009).



Table 7

*Revised Rubric for the First Four Tasks of the Education Research Project*

Research Question Index	1. Was the research question related to the identified problem?	0 = No 1 = Weakly Related 2 = Related 3 = Strongly Related
	2. Was the research question manageable?	0 = No (too broad or too narrow) 1 = Manageable
Search Index	#3. Did search terms cover all major and minor concepts in the research question?	0 = missing a major concept 1= all majors, but missing a minor, 2= all major and minor concepts.
	#4. Was there an appropriate choice of database?	0 = listed one or more non-usf databases 1 = all items listed are usf article databases
	#7. Were the essential elements present in the article citation of sufficient quality (general APA format) to enable another to find the article?	0= missing key elements 1= missing some elements 2=all essential elements present (author, year, title, journal, volume number)
Article Index	#5. Did student find article that addressed the research question?	0 = No 1 = Weakly Related 2 = Related 3 = Strongly Related
	#6 Did the student make an appropriate choice of an article?	0 = Inappropriate (not current, not peer-reviewed, or not empirical) 1= Appropriate

Data were also screened for multivariate outliers. Mahalanobis' distances for data points ranged from .540 to 14.0. There were no obvious indications of outliers to be removed from the data set. With regard to homogeneity of covariance matrices, Box's M

test was not statistically significant,  $\chi^2(6, N=52) = 10.6, p = .10$ ; thus there was no evidence that this assumption had been violated.

Table 8

*Descriptive Statistics for Sections A and B on the Three Indices for the Education Research Project.*

Index	Class	M	SD	Skewness	Kurtosis
Research Question	A	.412	1.30	-.712	-0.570
	B	-.481	2.00	-.151	-1.78
Search	A	.636	1.83	-1.55	1.80
	B	-.742	2.58	-.464	-1.19
Article	A	-.074	1.68	-2.00	3.63
	B	.086	1.47	-2.07	3.71

After the assumptions were deemed to have been tenable for the data set, a one-way MANOVA was conducted, and it indicated there was a statistically significant difference between the means for the two classes on the set of dependent variables, Wilks  $\Lambda = .792, F(3, 38) = 4.19, p < .05$ . This result can be interpreted as 21% of variability in the indices' scores coming from class membership. The multivariate effect size  $f^2$  was 0.26. This indicated a small effect size for class differences on this set of dependent variables. In summary, this analysis showed that the Level Two intervention was associated with significantly higher achievement on the set of three indices.

A serendipitous finding during the scoring of these first four tasks was evidence that some preservice teachers selected an article that was not related to their research question, and then went back and changed their research question to match the chosen article. These preservice teachers may have “settled” for an off-topic article simply to

complete the assignment quickly. This highlighted a limitation of the ERP to approximate the process of EBP. Based on this finding, it was determined that subsequent iterations would require preservice teachers to find three articles that addressed the research question. This change encouraged preservice teachers to develop their database search strategies until they were successful at finding articles on their chosen topics.

*Grounded theory analysis of ERP research application essays.* Whereas the global goal of the DBR agenda was to create an intervention that could prepare preservice teachers to apply research to practice, the goal of this avenue of inquiry was to document the “current understandings” (Cobb et al., 2003) of preservice teachers regarding the essay that described the application of education research to classroom teaching. Careful analysis of how preservice teachers in Group A (Level 2) communicated their intentions to apply research to their teaching was conducted for the purpose of understanding how to strengthen the effectiveness of the ERP. A full report of the results of this analysis can be found in van Ingen, McHatton, and Vomvoridi-Ivanovic (under review), and a succinct summary of the analysis is provided below.

Grounded theory was the method chosen for data analysis. Figure 5 provides an overview of the iterations of this data analysis. The primary researcher conducted line-by-line coding of the 29 documents. This resulted in 153 codes. Codes were analyzed, similar codes combined, and a codebook was created with the final 40 codes (see Appendix D). The documents were recoded using the 40 codes from the codebook. From constant comparative analysis of the coded documents, ten categories of codes emerged and then four broad themes emerged from these categories. Appendix E

provides a visual overview of the relationships between codes, categories, and themes.

<b>Fifth Iteration</b>			
From constant comparative analysis first categories of codes emerge and then four themes emerge:			
Understanding Research	Implementation Plan	Implementation Modification	Implementation Evaluation
<b>Fourth Iteration</b>			
Second researcher trained with the codebook. Second researcher coded 20% of documents. Consensus on application of the codes.			
<b>Third Iteration</b>			
Documents were recoded with the 40 defined codes.			
<b>Second iteration</b>			
Line-by-line open coding. 153 codes generated. Codes were analyzed and similar codes combined. A codebook was created with the resulting 40 codes.			
<b>First Iteration</b>			
Reading of 29 documents for holistic analysis			
Data	Data	Data	Data

*Figure 5.* Overview of the five iterations of data analysis for the application to research essays (Iteration 1).

The ten categories were as follows: (a) extent to which the research findings are identified, (b) stance toward research findings, (c) extent to which the plan is defined, (d) the plan’s connection to a stance on teaching and learning, (e) the plans’ connection to research, (f) the extent to which the plan considers local conditions, (g) the extent to which the plan identifies supports and barriers to implementation, (h) recognition of the need for iteration in plan implementation, (i) the plan to evaluate implementation, and (J) the plan to share results.

When the coded essays were evaluated in terms of the categories, a striking pattern was identified. There was a clear reduction in proficiency of preservice teacher skills from communicating understanding of the research, to creating an implementation

plan, to creating a plan for evaluating the implementation. Approximately one-third of the preservice teachers were able to communicate understanding of the research, about one-fourth created an effective plan to implement that research, and less than one-tenth planned to evaluate that implementation. Overall, these results indicated that, even after exposure to the ERP, preservice teachers were unable to communicate effective plans to apply research to future teaching. Specifically, these findings also suggested that preservice teachers may need greater support in creating and evaluating their research implementation plans than was provided in the ERP of Iteration 1.

*Treatment potency test.* In addition to the previously discussed avenues of inquiry, I designed a test of intervention potency. Given that the overarching research question for this DBR study was to understand how preservice teachers can be prepared to apply research to practice, I designed a proxy test to see if the preservice teachers, after experiencing the intervention, were interested in pursuing education research to inform their future teaching practice. At the end of the semester, participants from all three sections received an email, forwarded by the section instructors, that informed the preservice teachers of a Blackboard (an online learning platform) student organization site that offered free access to education research relevant to teaching elementary school mathematics. The email invited students to join this organization, by either emailing the PI or by signing up via their own Blackboard account (instructions on how to do this were included in the email).

After receiving the email offering access to math education research resources, only 1 student out of 33 students (3%) from Section A and one student out of 23 students (4%) from Section B signed up to have access to the Blackboard student organization

site. No student from Section C signed up for access to the site. This finding indicated that the intervention may not have been potent enough to have an effect that persisted beyond the semester's coursework.

**Iteration 2.** Iteration 2 was conducted in the summer of 2011 with one section of Mathematics Methods I. Although findings from the quasi-experimental study of Iteration 1 indicated benefit to collaboration with the research librarian, such collaboration was not feasible during this summer semester (the collaboration resumed in Iteration 3). Findings from the grounded theory analysis and the treatment potency test indicated the need to strengthen the impact of the ERP, particularly with regard to preservice teachers' ability to create and evaluate research implementation plans. Therefore, preservice teachers were required to incorporate their research findings into a micro-teaching lesson—a mini version of an elementary mathematics lesson that is taught to peers. In addition, throughout the course of the semester, students were asked to find five research articles on topics that followed the course syllabus (e.g., the teaching of fractions at the elementary school level). The purpose of this modification was to give preservice teachers more experience searching for articles as well as to provide them more exposure to the form of research articles.

Both at the beginning and the end of the semester, preservice teachers in Iteration Two completed a short version of the Information Literacy Questionnaire (Appendix F). This questionnaire has fourteen fewer items than the pilot version of the Information Literacy Questionnaire (Appendix B) that was used in Iteration 1.

At the end of the semester, preservice teachers in Iteration 2 were also given the same Treatment Potency Test that was used in Iteration 1. The result of this test, to see

whether preservice teachers would seek out access to education research, was that 4 preservice teachers out of 30 (13%) registered for access to the education research repository. This result showed an improvement over the one student per section in Iteration 1 (3% and 4%), but it still indicated lack of interest on the part of preservice teachers to extend their learning about EBP beyond the requirements of the course.

**Iteration 3.** Iteration 3 was conducted in the fall of 2011 with one section of Mathematics Methods II, the second of the two-semester sequence of mathematics methods courses for preservice elementary teachers. There were 34 students enrolled in this course. Eight of these students had completed an ERP in a previous Mathematics Methods I course. Based on the findings from Iteration 1 and the availability of the education research librarian, the two 75-minute collaborative workshops with the librarian were retained and revised to involve more active involvement of the preservice teachers. Between sessions with the librarian, preservice teachers completed a concept table (Figure 6) in which they identified major search terms. Online access to additional information literacy resources was also made available to the preservice teachers.

The micro-teaching requirement from Iteration 2 was retained and strengthened in the following manner. After completing the ERP, each preservice teacher consulted

## Information Research

NAME\_Math Teacher\_\_\_\_\_

1. Write a question for your topic. Be sure that it is sufficiently manageable for your research project.

**Do students learn better or worse using calculators in elementary math classes?**

2. Identify the essential concepts in your topic description and complete the keyword table below. Enter your essential concepts in the numbered Concept Categories cells below. Use less or more than five, as necessary. Enter your keywords (synonyms & related terms) for each concept in the columns below each concept category. **NOTE:** a concept term can also be a keyword

**KEYWORD TABLE**

Concept Categories	Students	Learning	Mathematics	Calculators	
Keywords (Synonyms & Related Terms)	Elementary School Students	Comprehension Retention Transfer	Computation		

3. Using terms from the keyword table above, write one or more search strategy statements for your topic. Use Boolean operators as needed.
4. Select appropriate database(s) for your topic. Write the name(s) of these databases below.

Figure 6. Concept table used to prepare for the Education Research Project



with a peer to create a plan to apply the research to a micro-teaching lesson. After the micro-teaching lesson, preservice teachers consulted with the same peer again, revised the lesson, and then taught the micro-teaching lesson again. After the second micro-teaching lesson, preservice teachers wrote reflections on the process applying research to practice. In these reflections, preservice teachers expressed that they felt prepared to apply research to practice in their future teaching careers. Some reflection excerpts are provided below:

When first discovering that I had to do research in this course, I thought that it didn't make sense and was actually quite irrelevant . . . After applying what I learned through the microteaching activity I can really say that I learned a lot from this research experience. I am walking away with knowledge that I most definitely did not have before and the confidence to teach word problem solving strategies to my future students. (Preservice Teacher 12, Fall 2011)

I have actually enjoyed applying research to practice; at first I was hesitant and thought it was a little irrelevant however I do see the importance. Researching helped me become more aware of the subject matter. I got to see what I researched from multiple perspectives and it helped me with my lesson. (Preservice Teacher 5, Fall 201)

When we first began our research project, I was convinced that it would not help me in anyway, and that I would never apply what I learned. However, since you had us apply our research in our micro-teachings, I came to realize that I will most

definitely use my research in my classroom. Also, I feel as if researching more math topics will come in handy in the classroom. If I had any problems with say, an ELL or a student with a disability, I could go and find a study on that topic, and then apply what the material said in my classroom.

(Preservice Teacher 2, Fall 2011)

Before beginning this project, I felt overwhelmed by the prospect of doing research and adapting the work of other researchers to my lesson plan. I had this preconceived notion that research was divorced from the real-world of teaching and that it would be too time-consuming to carry out in a real world situation. Now that I have completed this process of researching a topic and put it into practice, I see that the research that I did enables me to learn to plan and teach more effectively. For my research topic in particular, it really allowed me to teach my lesson in a way that improved my “students” understanding of the subject matter. I am now sold on the idea of applying research to practice and plan to continue regularly incorporating research into my lesson planning.

(Preservice Teacher 17, Fall 2011)

These reflections provided anecdotal information on the climate variable of engagement and the learning variable of disposition that Collins et al. (2004) recommended for DBR studies. In summary, these reflections presented the preservice teachers’ perspectives on the benefit of utilizing the micro-teaching and the overall success of the intervention. Clearly, there were preservice teachers who understood and

valued the process of applying research to practice. The treatment potency test was also conducted in Iteration 3 and resulted in 6 out of 25 (25%) of the teachers signing up for access to the resource for elementary level mathematics education research.

After Iteration 3 was completed, I interviewed three students—one from each of the first three iterations. The purpose of the interviews was to gather information on how these preservice teachers perceived the impact of the ERP on their internship teaching experiences a semester or two semesters after completion of the project. The protocol for these interviews is in Appendix G. Interview participants were recruited by an email that was sent to all of the participants in Iterations 1, 2, and 3. Interview participants did not receive compensation for their participation in the interviews. Interviews lasted 23 minutes, 18 minutes and 38 minutes respectively. Each interview was transcribed and open-coding was used to analyze the content of the interviews. Although the three interviewees each talked about the importance of the ERP, none of the three had engaged in the process of applying research to practice during their internships. The first two interviewees reported no thoughts about applying research to practice while engaged in their internship teaching. The third interviewee reported thinking about the research from the ERP but did not implement the research in classroom teaching. This finding—that the preservice teachers valued the ERP but did not engage in the process of EBP during their internships— influenced the design of Iteration Five in which preservice teachers were given the opportunity to apply research to practice during internship teaching.

**Iteration 4.** Iteration 4 was conducted in the summer of 2012 with one section of Mathematics Methods I. The design of the intervention for Iteration 4 consisted of the Revised ERP (Table 9) as well as the two micro-teaching experiences

Table 9

*Revised Education Research Project Used in Iteration Four*

Education Research Project Tasks	
1. Question Formulation	1. Identify a classroom-based problem for teaching mathematics.
	2. Turn this problem into a general question.
	3. Turn this general question into a question that can be researched.
2. Evidence Search	4. List key words and descriptors that you will use for your literature search.
	5. List three databases that you will use for your search.
	6. Choose three articles that relate to your question. At least two must be scholarly, empirical journal articles. The other one may be scholarly, theoretical or practitioner articles. All must be peer-reviewed.
	7. Create an annotated bibliography a. APA citation
3. Critically appraise research	7b. 6-12 sentence summary of essential points of each article
	8. Research synthesis: Based on these three articles, how do you now understand your initial problem? (Approx 1 double-spaced page)
4. Implement the intervention	9. Create a written plan for how you will apply this research to your classroom teaching. Consult with a peer on this plan. In the written plan: clearly connect the research implementation to your stance on teaching and learning (the purpose for your actions), clearly define your plan, clearly connect your plan to the research you read, consider your local conditions, discuss any barriers to the implementation, discuss any supports that you can leverage for the implementation, provide a clear plan as to how you will evaluate the implementation.

that were first used in Iteration 3. The micro-teaching experiences required preservice teachers to apply their research findings to mini-lessons that they taught to their peers.

Unlike Iterations 2 and 3, the intervention in Iteration 4 did not require preservice teachers to find five additional research articles throughout the course of the semester. The decision to eliminate this requirement did not come from evidence of its lack of importance but rather from the fact that the summer term is shorter than the fall and spring terms. There was not enough time in the course to include this series of tasks in the intervention.

In Iteration 4, I piloted a rubric to evaluate the ERP (Appendix H). There was a total of 25 points possible, and observed scores ranged from 20.50 to 25.00 ( $M = 24.29$ ,  $SD = 1.30$ ). In addition, I created a summative assessment, described in the following passage, to evaluate preservice teachers' ability to transfer their EBP skills to novel situations.

At the completion of the semester, I asked preservice teachers to read an excerpt from an Institute for Educational Sciences (IES) practice guide titled "Developing Effective Fractions Instruction for Kindergarten Through 8<sup>th</sup> Grade" (Siegler et al., 2010) that had been published on the What Works Clearinghouse website (see Appendix I for the excerpt). I asked preservice teachers to describe through a brief essay how they would apply the research from the summary to their classroom teaching. The purpose in giving this summative assessment was to obtain evidence of preservice teachers' abilities to engage in an approximation of the process of EBP. In Iteration 5 I created a rubric to evaluate the summative assessment (Appendix J). The Treatment Potency Test for Iteration 4 resulted in 6 out of the 30 students (20%) registering for access to additional educational research.

**Iteration 5.** This section provides information on the following aspects of Iteration 5 of the DBR process: participants, intervention, ecological environment, system of assessment, instrumentation, data collection, and data analysis.

**Participants.** Participants in the first four iterations on the DBR study were preservice elementary education majors enrolled in a traditional elementary education program. For the fifth iteration, the 12 participants were enrolled in a residency version of the elementary education program at the same large, southeastern university. I refer to these participants as the *residents* through the rest of this document. During all four semesters of the residency program, the residents were in a K-5 classroom during the majority of the school day. During semesters three and four, the residents assumed responsibility for a significant portion of classroom instruction. The residents in this study were in their third semester in the program and enrolled in Mathematics Methods I. All twelve residents were female. One fourth of the residents self-identified as a racial minority. Eleven of the twelve residents were traditional college students in their twenties. One of the residents was a non-traditional student who had already had a non-educational career.

**Intervention.** As was the case for each of the four previous iterations, the Education Research Project (ERP) was the central component of the intervention in Iteration Five. The first nine tasks of the ERP were identical to the Revised ERP that was used in Iteration Four (Table 9). In addition to these nine tasks, two additional tasks were added to the ERP-Residency Version (see Table 10). The additional tasks ten and eleven required the residents to apply their research findings to their classroom teaching and

Table 10

*Education Research Project- Residency Version Used in Iteration Five*

Education Research Project Tasks	
1. Question Formulation	1. Identify a classroom-based problem for teaching mathematics.
	2. Turn this problem into a general question.
	3. Turn this general question into a question that can be researched.
2. Evidence Search	4. List key words and descriptors that you will use for your literature search.
	5. List three databases that you will use for your search.
	6. Choose three articles that relate to your question. At least two must be scholarly, empirical journal articles. The other one may be a scholarly, theoretical or practitioner article. All must be peer-reviewed.
	7. Create an annotated bibliography a. APA citation
3. Critically appraise research	7b. 6-12 sentence summary of essential points of each article
	8. Research synthesis: Based on these three articles, how do you now understand your initial problem? (Approx 1 double-spaced page)
4. Implement the intervention	9. Create a written plan for how you will apply this research to your classroom teaching. Consult with a peer on this plan. In the written plan: clearly connect the research implementation to your stance on teaching and learning (the purpose for your actions), clearly define your plan, clearly connect your plan to the research you read, consider your local conditions, discuss any barriers to the implementation, discuss any supports that you can leverage for the implementation, provide a clear plan as to how you will evaluate the implementation.
	10. Video-tape a lesson in which you apply the research to your teaching.
5. Evaluate the effect on teaching and learning	11. You will write three reflections that are meant to accompany the video. <ul style="list-style-type: none"> <li>a. Reflection on Teaching Actions: Here you briefly describe what you did to implement the research. Explain how your actions were connected to the research that you read. How did you make adjustments to the research to meet the needs of your unique classroom? How did you evaluate the implementation?</li> <li>b. Reflection on Implementation Impact: What was the effect of the implementation on the students and you the teacher? Discuss how you perceived the effectiveness of the application of research to support your students' learning. Give concrete examples of the success or lack of success based on the evidence from the video or documents.</li> <li>c. Reflection on Future Plans: Will you revise the implementation? Will you continue using it, or will you stop implementation and <u>why</u>?</li> </ul>

then to reflect upon that evaluation. Residents were required to create a portfolio that contained video and/or documentary evidence of how they applied research to their classroom teaching and complete a set of three reflections. The requirement to apply the research to classroom teaching was a closer approximation to engagement in the process of EBP than the use of micro-teaching in Iterations 2, 3, and 4.

**Ecological environment.** The previous section provided an overview of *what* the intervention for Iteration 5 required. This section provides an overview of *how* the intervention was implemented. During the first weeks of the fall 2012 semester, residents were given access to online videos and websites that provided them with basic information literacy skills (e.g., how to access the university's online library materials, how to use ERIC thesaurus). Residents were asked to view these resources and complete a brief quiz (Appendix L) to demonstrate understanding of the material. Created by the education librarian, the quiz was composed of eleven multiple choice, dichotomously scored items. Residents were required to complete the quiz, but the quiz was not used to calculate course grade. Observed scores ( $n=12$ ) ranged from 8 to 11 ( $M=10.33$ ,  $SD=0.89$ ). Cronbach's *alpha*, a measure of internal reliability, was .75. Results were reviewed by the course instructor and librarian before further instruction.

Following completion of the quiz, the course instructor co-taught a class with the university's education librarian on how teachers can transform their classroom based questions and problems into questions that are answerable through searches of education research. Part of this class involved a workshop where residents identified their own classroom-based problems and practiced turning these problems into questions that could be answered by education research. Residents were then given homework time to



complete concept tables (Figure 5, p. 102). Several days later, residents met at the library for a second co-taught workshop where residents learned how to search online databases for research literature that addressed their questions. Each resident selected three articles that addressed the identified research question.

Over the course of two months, each resident completed independently tasks one through nine (referred to as Part I) of the ERP-Residency Version (Table 8). At the end of the two months, residents were required to submit their written documentation of Part I (tasks 1-9) of the ERP-Residency Version. Afterwards, residents were asked to work on Part II (tasks 10-11) of the ERP-Residency Version. They implemented their research findings in their classroom teaching and documented this implementation through video and/or artifact evidence. Following this documentation, they wrote brief reflections on: 1) what they did in the classroom, 2) the impact that the implementation had on the students and the teacher, and 3) their future plans to continue implementation, revise implementation or end the implementation and why they made this decision.

**System of assessment.** There were two objectives to this study: 1) to create an intervention that equips preservice teachers to engage in the process of EBP and 2) to create a system of assessment to measure the extent to which preservice teachers can engage in the process of EBP. This section describes the system of assessment that accompanied the intervention. My decision to create a *system* of assessment was in direct response to Hill and colleagues' (2012) call for such systems. They argued that “major instrument developers—including states, researchers, and other nongovernmental entities—must go beyond simply writing instruments; they must create observational

*systems* in which quality observational instruments, well-trained raters, and robust scoring designs are combined to produce reliable teacher scores” (p. 56).

The *process* of EBP is, by definition, a performance in which teachers engage. In designing a system for a summative performance assessment for this process, I also looked to the work of Boerst, Ball, and colleagues at the University of Michigan (Boerst et al., 2012). They have developed *standardized* performance assessments in which the context of the assessment has been simplified so that all respondents experience the same set of *standardized* conditions. I have chosen to develop a *standardized* performance assessment to measure a resident’s ability to engage in the last two steps of the EBP process. Residents were given the opportunity to gain familiarity with the format of the standardized performance assessment through the administration of a practice assessment during the week prior to the actual assessment. During the practice, residents were given an additional research excerpt and explicitly informed of the content required for a successful performance.

In the week following the practice, residents were given two elementary school level research synopses. Each synopsis came from an Institute for Educational Sciences (IES) practice guide (Siegler et al., 2010; Woodward et al., 2012) that had been published on the What Works Clearinghouse website (See Appendix I for the research excerpts). Residents read the brief reports and then wrote one essay for each synopsis in which they described how they would apply that research to their mathematics teaching. These essays were scored with a rubric (Appendix J) by three raters. A full description of the scoring process can be found in the *data analysis* section. Use of Generalizability

Theory allowed me to provide information on the reliability of these summative performance assessment scores given the conditions of two tasks and three raters.

In an effort to triangulate data and to explore if writing style or writing fluency influenced test performance, I invited residents on a volunteer basis to engage in a brief interview in which I verbally questioned the residents about their written responses (see Appendix P for the interview protocol). I used open-ended probes to inquire if students knew more or less than their writing indicated. The verbal responses were compared to the written responses; similarities and discrepancies were documented.

**Instrumentation.** There were four instruments used to gather data in the Fifth Iteration. Each instrument and its use are briefly described below.

*1) The Information Literacy Questionnaire- Short Form* (Appendix F). The purpose of this questionnaire was to gather data on 1) participant demographics and history related to teaching and information literacy, 2) intentions to use research to inform future teaching practice, and 3) information literacy skills. The full version of this instrument was piloted in Iteration 1 and revised for Iteration 2. The instrument has been used in each iteration and allows for comparison across iterations. In Iteration 5, this instrument was given three times: twice for pretest-posttest comparison, once at the beginning and once at the end of the semester, and a third time one week after posttest to assess for test-retest reliability. Because additional psychometrics of this questionnaire (e.g., underlying factor structure, internal reliability) have not been evaluated, only pretest-posttest change on individual items was analyzed. Again, individual item reliability was calculated by comparing the second and third scorings.

- 2) *The Familiarity with the Process of Evidence-Based Practice in Education Scale* (Appendix K). The purpose for this instrument was to measure residents' perception of their familiarity with the process of applying research to practice. This scale was modeled after a similar scale that has been used in the field of social work (Parish & Rubin, 2011; Rubin & Parrish, 2009, 2010). The Familiarity scale has been reviewed by these experts: Sarah Bleiler (mathematics educator with instructional experience with the ERP), Robert Dedrick (expert in measurement), and Danielle Parrish (professor of social work and expert in EBP). Appendix K contains a summary of the completed review forms that were used by the reviewers. The Familiarity with EBP in Education scale was piloted in this iteration for future development. This scale was given at the beginning and the end of the semester. At the end of the semester, the scale was given twice—one week apart—in order to assess test-retest reliability.
- 3) *Education Research Project (ERP)-Residency Version* (Table 10). The ERP-Residency Version was the intervention itself in Iteration 5, but it was also a platform for collecting data about residents' abilities to engage in the process of EBP. For each of the 11 tasks in the ERP-Residency Version, residents submitted documentation of their completion of the task. This documentation was analyzed to provide information about the extent to which the residents were successful in applying research to practice.
- 4) *Standardized Performance Assessment*. The purpose of this assessment was to measure the extent to which residents were able to describe how they intended to engage in the process of applying research to practice. Residents were asked to

read a research summary (e.g., Siegler et al., 2010; Woodward et al., 2012) and then to write an essay about how they would apply that research to their teaching practices. Although the standardized performance assessment was created to be a summative assessment, residents were asked to complete this standardized performance assessment both at the beginning and the end of the semester. This allowed for evaluation of the extent to which the performance assessment was sensitive to the intervention discussed previously.

**Data collection.** In the first week of the semester, residents completed the Information Literacy Questionnaire-Short Version, the Familiarity with the Process of Evidence-Based Practice in Education scale, and submitted two essays for a pretest Standardized Performance Assessment. Each of these measures was taken on a different day of the first week of the semester. During the semester, residents worked on the Education Research Project- Residency Version Part I (Tasks 1-9) for eight weeks. This was submitted to the instructor and then residents worked on Part II (Tasks 10-11) for the next six weeks. During the last two weeks of the semester, residents again completed the Information Literacy Questionnaire-Short Version, the Familiarity with the Process of Evidence-Based Practice in Education scale, and submitted two essays for the Standardized Performance Assessment. Again, each of these assessments was completed on different days.

**Data analysis.** Consistent with DBR, I analyzed data from multiple sources (e.g. the ERP documents, the Information Literacy Assessment, the Familiarity with EBP scale, and the standardized performance assessments) in order to describe the extent to which the residents were able to apply research to practice. In Chapter One, I outlined

two products that would result from this study: a teaching intervention that mathematics teacher educators can use in order to prepare preservice teachers to engage in the process of evidence-based practice (a routine of practice) and a system of performance assessment that measures this practice. I now describe the data analysis procedures that corresponded to the development of each of these two products.

*The teaching intervention.* The purpose for the data analysis related to the teaching intervention was to examine the extent to which the intervention prepared the residents to engage in the process of evidence-based practice. I looked to the data collected from the first three instruments listed previously (The Information Literacy Questionnaire-Short Version, The Familiarity with the Process of Evidence-Based Practice in Education Scale, and Education Research Project-Residency Version (ERP-RV) to describe the extent of the effectiveness of the teaching intervention.

Because the items on the Information Literacy Questionnaire Short Form (Appendix F) were used in each iteration, I looked at the pre-post changes in Iteration 5 and compared these changes to the changes seen in the previous iterations. Although I did not use total score on this questionnaire, I looked at change from pretest to posttest on individual items (Item numbers from the pilot version were renumbered to match the short form item numbers). First I examined the quantitative data from Iteration 5 to see if the assumptions of normality and homogeneity of variance were tenable for items with Likert-type response scales that were considered to be continuous. I provided descriptive data (i.e., means and standard deviations) for the change scores of these items. I also provided odds ratios for items that were dichotomously scored. Due to anticipated lack of power (small sample size) and the need for family-wise Type I error control, I

conducted only two *t*-tests on the pre-post data from Iteration 5. I conducted a dependent *t*-test on the pre and post scores for item 6d and another dependent *t*-test on the pre and post scores for item 6e. I chose those two items because they addressed resources, scholarly and practitioner journal articles, that were specifically targeted by the intervention. Then I conducted an ANOVA on the change scores for items 6d and 6e for all five iterations. I used a modified Bonferroni approach to Type One error rate control with the ANOVA tests. I also conducted two independent means *t*-tests to compare change scores on Items 6D and 6E for participants in Iteration 1 and those in Iteration 5.

For the total scores on the Familiarity with Evidence-Based Practice Scale, I determined if the data were normally distributed and if there was homogeneity of variance. Without evidence that these conditions were violated, I conducted the dependent means *t*-test. Because both the *N* was small and previous iterations have shown medium to small effects for the teaching intervention, I anticipated that the power for the *t*-test may be very low. As an alternative way to evaluate change in the individual participant from pretest to posttest, I also calculated reliable change index (RCI) scores for each participant (Jacobson & Traux, 1991; Wise, 2004). The RCI allows a researcher to look to see if there is clinically significant change. I used the Jacobson, Follet, and Revensdorf (1984) RCI equation:  $RCI = \frac{\bar{x}_1 - \bar{x}_2}{SE}$  where  $SE = s_1 \sqrt{1 - r_{xx}}$  and  $s_1$  is the standard deviation of the pretreatment group and  $r_{xx}$  is the test-retest reliability. Because cut-off scores for reliable change in the context of EBP training have not been established, it was not possible to categorize resident scores as clinically relevant. Nevertheless, it was possible to look at the spread in RCI scores.

I computed descriptive data on the scores for the Education Research Project (see Appendix H for the scoring rubric). I then used the codebook created from Iteration 1 (Appendix D) to code the application essays from this project. I compared the results of this coding with the results of the coding from the first iteration. This analysis allowed me to compare the effect of the more rigorous version of the learning module from the fifth iteration to the less rigorous version in the first iteration.

*The system of assessment.* In this section I describe the procedures for analyzing the data from the Standardized Performance Assessment. In response to Hill et al. (2012), I provide information on the entire system involved in the assessment. First I describe the scoring procedures for this assessment. I recruited and trained two additional teacher educators beside myself to be raters for the Standardized Performance Assessment. We used performance assessments from Iteration Four for training purposes. The training required raters to become familiar with the rubric in Appendix J. After discussion of each level of each indicator on the rubric (examples for each level of the indicator provided from documents from Iteration Four), each rater independently rated three of the Iteration Four essays. The three raters then compared scores and discussed discrepancies. Raters again rated three more essays independently and then compared scores and discussed discrepancies. Inter-rater reliability for the second round of scoring was calculated through the following procedure. The scored rubrics for the three essays were compared at the item level. Because there were 3 scored essays and 7 items on the rubric, there were a total of 21 item scores that were compared. Agreement of all three raters on an item was coded as a one. Disagreement by any rater was coded as a zero.



The reliability codes (ones and zeros) were summed and divided by 21 (the total number of items). Training cycles continued until inter-rater reliability was greater than 80%.

After training, the three raters rated the two posttest standardized performance assessment essays from Iteration Five. Each rater scored each essay twice with a minimum of one day between scoring occasions. Scores were recorded on separate spreadsheets for each occasion.

Following the recommendation of Hill et al. (2011), I used Generalizability Theory to analyze the data from the Standardized Performance Assessments. As was discussed in greater detail in Chapter Two, in a G study, the observed variance in total score is partitioned through analysis of variance (ANOVA) and estimates of variance components are obtained. In other words, it allows the researcher to determine how much of the score variance is due to true variance from the person (the test taker) and how much is due to error from testing conditions such as rater, occasion, or task. These variance estimates are used to generate a generalizability coefficient that is similar to a reliability coefficient in Classical Test Theory. The formula for the generalizability coefficient is  $\frac{\hat{\sigma}_p^2}{\hat{\sigma}_p^2 + error}$  where  $\hat{\sigma}_p^2$  is the variance due to person (true variability) and error is the variance due to error. Generalizability coefficients can be calculated in order to make a relative decision (the reliability of being able to order the scores) or to make an absolute decision (the reliability of the scores). I calculated both relative and absolute generalizability coefficients in order to have maximum amount of information about the reliability of this system of assessment. The formula that I used to calculate the relative error variance was:

$$\text{Relative error variance} = \frac{\hat{\sigma}_{pr}^2}{n_r} + \frac{\hat{\sigma}_{pt}^2}{n_t} + \frac{\hat{\sigma}_{po}^2}{n_o} + \frac{\hat{\sigma}_{prt}^2}{n_r n_t} + \frac{\hat{\sigma}_{prv}^2}{n_r n_o} + \frac{\hat{\sigma}_{ptv}^2}{n_t n_o} + \frac{\hat{\sigma}_{prtv}^2}{n_r n_t n_o}.$$

The formula that I used to calculate the absolute error variance was:

Absolute error variance =

$$\frac{\hat{\sigma}_r^2}{n_r} + \frac{\hat{\sigma}_t^2}{n_t} + \frac{\hat{\sigma}_o^2}{n_o} + \frac{\hat{\sigma}_{rt}^2}{n_r n_t} + \frac{\hat{\sigma}_{to}^2}{n_t n_o} + \frac{\hat{\sigma}_{ro}^2}{n_r n_o} + \frac{\hat{\sigma}_{rtv}^2}{n_r n_t n_o} + \frac{\hat{\sigma}_{rv}^2}{n_r n_o} + \frac{\hat{\sigma}_{tv}^2}{n_t n_o} + \frac{\hat{\sigma}_{pr}^2}{n_r} + \frac{\hat{\sigma}_{pt}^2}{n_t} + \frac{\hat{\sigma}_{po}^2}{n_o} + \frac{\hat{\sigma}_{prt}^2}{n_r n_t} + \frac{\hat{\sigma}_{prv}^2}{n_r n_o} + \frac{\hat{\sigma}_{ptv}^2}{n_t n_o} + \frac{\hat{\sigma}_{prtv}^2}{n_r n_t n_o}.$$

In this analysis I had a fully-crossed 3 facet G-study design. The facets were rater, task, and occasion. The design was fully-crossed because: each person was scored by each rater, each person completed the same two tasks, and each rater scored each task twice. Appendix N provides a table template that I used to report the variance components. Appendix O shows how the data was set up in order to conduct the G-studies.

Once the G-studies were completed, I conducted a series of decision (D) studies. D-studies allow the researcher to identify the testing conditions that would result in a given level of reliability. I identified the testing conditions in my study that would lead to generalizability coefficients of: 1) 0.7 and 2) 0.8. For both of these generalizability coefficient values, I conducted D-studies for an absolute decision and for a relative decision. In the end, the purpose of the D-studies was to enable me to make recommendations about the conditions that maximize the reliability of this standardized performance assessment.

To determine the extent to which the standardized performance assessment was sensitive to instruction, one rater coded both pretest standardized performance assessments twice and both posttest standardized performance essays twice. The means

of the average pretest scores and the average posttest scores were compared through a *t*-test.

**Summary.** The overall goal of the study was to identify how to prepare preservice teachers to engage in the process of applying research to practice. The two objectives of the study were to produce two products: a teaching intervention that can be used by teacher educators and a system of assessment that supports that intervention. Each iteration has been designed to produce empirical knowledge that can contribute to these objectives, and the knowledge gained from each iteration was used to refine and further develop the products.

There were four specific research questions, introduced in Chapter 1, that guided the design of the fifth iteration. These questions were:

- 1) After experiencing the intervention, to what extent did the participating preservice teachers demonstrate the ability to apply research to their classroom teaching?
- 2) To what extent did preservice teachers' intentions to apply research to future teaching change from the beginning to the end of the intervention?
- 3) What, if any, were the differences in a) ability to apply research to practice and b) intention to apply research to practice between participants in the fifth iteration and participants in previous iterations?
- 4) How reliable was the system of assessment that was used to assess preservice teachers' abilities to apply research to teaching?

Question one was addressed through the analysis of the data collected from the ERPs (Table 10), the Summative Performance Assessment, and the Familiarity with EBP in Education Scale. Question two was answered through analysis of pretest posttest changes

on the Information Literacy Questionnaire. Question three was addressed through the comparison of the coded application to research essays (from the ERPs) from Iteration One and Iteration Five, and the comparison of pretest posttest change scores from items on the Information Literacy Questionnaire for all five iterations. Question four was answered through the Generalizability study of the system of standardized performance assessment.

## Chapter 4: Results

In this chapter I present the results of the analyses that were conducted on the following sources of data from the fifth iteration of this study: 1) the Information Literacy Questionnaire-Short Version, 2) the Familiarity with the Process of Evidence-Based Practice in Education Scale, 3) the Education Research Project (ERP)-Residency Version, and 4) the Standardized Performance Assessment. In Chapter 5, I interpret the results and discuss research and practice implications.

### **The Information Literacy Questionnaire-Short Version**

The residents from Iteration 5 completed the Information Literacy Questionnaire-Short Version (Appendix F) three times: once during the first week of the semester and twice, separated by an interval of one week, during the last two weeks of the semester. The last two posttest administrations were used to calculate test retest reliability coefficients at the item level. The item reliabilities ranged from .40 to 1.0 ( $M = .74$ ,  $SD = .19$ ) and are presented in Table 11. Seven of the 15 items had reliabilities less than .70. These low reliability levels mean that pretest-posttest comparisons of these items ought to be interpreted with caution.

**Descriptive statistics.** Tables 12 – 15 provide descriptive statistics on the pretest-posttest change at the item level for Iteration 5. For the purpose of being able to compare Iteration 5 changes to those of previous iterations, these tables include the descriptive statistics for the first four iterations as well. Tables 12, 14, and 15 provide means, standard deviations, and Cohen's  $d$  effect sizes for the questionnaire items that have

Table 11

*Test Retest Reliability Coefficients for the Information Literacy Questionnaire Items*

Item	Reliability Coefficient
1. How many times here at USF has a librarian visited one of your classes, OR have you been to the library to receive group instruction on the use of the library and its resources?	.90 <sup>a</sup>
2. How many times have you communicated one-on-one with a librarian to get assistance in an information search (face-to-face, email, or chat)?	.61 <sup>a</sup>
3. Have you ever been required to find scholarly articles as references for a required paper/project in an EDUCATION course here at USF?	1 <sup>c</sup>
4. How much experience do you think that you have with READING education research?	.76 <sup>a</sup>
5. As far as you know, are teachers required by law to use teaching practices that are supported by education research?	.82 <sup>b</sup>
6A. When you think about your future teaching career, how likely do you think you are to use advice from fellow teachers to help inform your teaching?	.70 <sup>a</sup>
6B. When you think about your future teaching career, how likely do you think you are to use advice from administration to help inform your teaching?	.69 <sup>a</sup>
6C. When you think about your future teaching career, how likely do you think you are to use Google searches to help inform your teaching?	.82 <sup>a</sup>
6D. When you think about your future teaching career, how likely do you think you are to use scholarly research journal articles to help inform your teaching?	.45 <sup>a</sup>
6E. When you think about your future teaching career, how likely do you think you are to use practitioner journal articles to help inform your teaching?	.62 <sup>a</sup>
6F. When you think about your future teaching career, how likely do you think you are to use textbooks from USF education courses to help inform your teaching?	.40 <sup>a</sup>
7. What is the purpose of an abstract?	.62 <sup>b</sup>
8. A periodical article has a methods and a results section, it is most likely a(n)	.82 <sup>d</sup>
9. Which of the following is characteristic of practitioner articles?	1 <sup>c</sup>
10. Which of the following questions could be answered by a research study carried out in an elementary school?	1 <sup>c</sup>
11. Let's say that you want to try out a teaching strategy that has been shown to be highly effective in a research study. The research article tells you about the strategy, but you have to decide how to apply that strategy in your own classroom. Circle any of the following factors that you would take into consideration when thinking about how to apply the research to your practice:	.57 <sup>a</sup>

*Note.* <sup>a</sup>Intraclass Correlation Coefficient; <sup>b</sup>Kappa Coefficient; <sup>c</sup>All pretest and posttest scores identical; <sup>d</sup>No variance in pretest, 9 of 11 posttest scores identical to pretest scores.

likert-type response scales. Table 13 provides odds ratios (odds of correct response/odds of incorrect response) and  $d_{\text{Cox}}$  values, an effect size for dichotomous variables that is analogous to Cohen's  $d$  (Cox, 1970; Haddock, Rindskopf, & Shadish, 1998; Sanchez-Meca, Chacon-Moscoso, & Marin-Martinez, 2003), for the questionnaire items that have dichotomous response scales.

Table 12 reports on two items related to librarian instruction. Because participants in Iterations 1C and 2 did not experience the librarian workshops, one would expect that these change scores (Items 1 and 2) would be lower than those of the other iterations. This is in fact the pattern that is seen. Also, of note on this table is the increase in change score for Item 4 (experience reading education research) from the first iteration to the fifth. The effect sizes for the revised iterations (2-5) are larger than those seen in the first iteration. Table 13 reports on change scores for multiple-choice items (items 7-10) that tested information literacy skills. Unlike Table 12, there is no discernable pattern seen in these change scores.

Tables 14 and 15 report on the questionnaire items 6A-6F that asked participants to indicate their likelihood to use various resources to inform their classroom teaching. Table 14 reports the data from the items (6A, 6B, 6C, and 6F) that ask about resources that were not specifically targeted during the ERP intervention. There is no apparent pattern to the change scores reported in this table. Table 15 reports the data from the items (6D and 6E) that were addressed through the ERP intervention. The effect sizes for the fifth iteration are, as expected, larger than for the previous iterations.

Table 12

*Changes in Pretest and Posttest Scores on Information Literacy Questionnaire Items 1, 2, 4, and 11 for Iterations 1-5*

	1. Group Instruction from a Librarian <sup>a</sup>			2. 1-on-1 Librarian Assistance for an Information Search <sup>a</sup>			4. Experience Reading Education Research <sup>b</sup>			11. Factors for Consideration when Implementing Research		
	$\Delta M$	Pooled <i>SD</i>	Cohen's <i>d</i>	$\Delta M$	Pooled <i>SD</i>	Cohen's <i>d</i>	$\Delta M$	Pooled <i>SD</i>	Cohen's <i>d</i>	$\Delta M$	Pooled <i>SD</i>	Cohen's <i>d</i>
Iteration 1A	1.00	0.95	1.05	0.35	1.00	0.35	0.10	0.38	0.25	0.29	1.50	0.19
Iteration 1B	0.04	1.06	0.04	0.13	0.93	0.13	0.00	0.42	0.00	0.21	1.57	0.13
Iteration 1C <sup>d</sup>	0.41	0.85	0.49	0.06	0.94	0.06	0.00	0.50	0.00	0.29	1.58	0.19
Iteration 2	0.37	0.97	0.38	-0.20	1.09	-0.18	0.27	0.47	0.57	0.77	1.57	0.49
Iteration 3	0.92	0.85	1.08	0.04	1.03	0.04	0.30	0.58	0.52	0.63	1.63	0.38
Iteration 4	1.11	0.91	1.23	0.07	0.77	0.10	0.37	0.50	0.74	0.30	1.34	0.22
Iteration 5	0.75	0.87	0.86	0.50	1.04	0.48	0.50	0.44	1.13	0.42	1.68	0.25

*Note.* Pooled *SD* = pretest posttest pooled standard deviation; Cohen's *d* = measure of effect size.

<sup>a</sup>Response scale from 0 to 3, <sup>b</sup>Response scale from 0 to 2, <sup>c</sup>Response scale from 0 to 7

<sup>d</sup>=Control group did not experience the Education research project.



Table 13

*Changes in Pretest and Posttest Scores on Items 3, 5, 7, 8, 9, and 10 for Iterations 1-5*

	3. Requirement to Use Ed Research		5. Does Law Require Research Supported Practices		7. Purpose of Abstract		8. Characteristics of Empirical Articles		9. Characteristics of Practitioner Articles		10. Identification of Question for a School-Based Study	
	OR	$d_{Cox}$	OR	$d_{Cox}$	OR	$d_{Cox}$	OR	$d_{Cox}$	OR	$d_{Cox}$	OR	$d_{Cox}$
Iteration 1A	20.08	0.79	0.57	-0.15	1.00	0.00	30.45	0.90	0.68	-0.10	1.40	0.09
Iteration 1B	57.91 <sup>a</sup>	1.07	1.00	0.00	4.60	0.40	1.40	0.09	0.47	-0.20	1.18	0.04
Iteration 1C <sup>b</sup>	1.28	0.07	0.19	-0.44	1.35	0.08	0.61	-0.13	1.68	0.14	1.44	0.10
Iteration 2	29.00	0.89	0.19	-0.44	7.00	0.51	61.00 <sup>a</sup>	1.08	2.68	0.26	1.00	0.00
Iteration 3	12.89 <sup>a</sup>	0.67	0.42	-0.23	7.00 <sup>a</sup>	0.51	9.47	0.59	11.50	0.64	1.24	0.06
Iteration 4	15.63	0.72	0.77	-0.07	37.81 <sup>a</sup>	0.96	51.07 <sup>a</sup>	1.04	0.68	-0.10	1.00	0.00
Iteration 5	8.33 <sup>a</sup>	0.56	0.71	-0.09	17.86 <sup>a</sup>	0.76	11.00	0.63	1.33	0.08	3.67	0.34

*Note.* All items dichotomously scored.

OR = odds ratio of posttest odds over pretest odds.  $d_{Cox} = L_{OR}/1.65$  is a measure of effect size

<sup>a</sup>-.5 was added to the numerator and denominator of the posttest odds due to the fact that the denominator was equal to zero.

<sup>b</sup>=Control group did not experience the Education research project.

Table 14

*Changes in Pretest and Posttest Scores on Items 6A, 6B, 6C, and 6F for Iterations 1-5*

	Advice from Teachers			Advice from Administration			Google Search			Course Textbooks		
	$\Delta M$	Pooled <i>SD</i>	Cohen's <i>d</i>	$\Delta M$	Pooled <i>SD</i>	Cohen's <i>d</i>	$\Delta M$	Pooled <i>SD</i>	Cohen's <i>d</i>	$\Delta M$	Pooled <i>SD</i>	Cohen's <i>d</i>
Iteration 1A	-0.06	0.60	-0.11	-0.13	0.55	-0.23	0.16	1.10	0.15	-0.03	0.88	-0.04
Iteration 1B	0.00	0.60	0.00	0.17	0.65	0.25	0.38	1.22	0.31	0.08	1.00	0.08
Iteration 1C <sup>a</sup>	0.06	0.65	0.09	0.41	1.13	0.36	-0.06	0.98	-0.06	-0.06	1.26	-0.05
Iteration 2	0.07	0.65	0.10	0.03	0.73	0.05	-0.17	1.13	-0.15	0.13	1.10	0.12
Iteration 3	-0.21	0.60	-0.35	-0.21	0.77	-0.27	0.04	1.13	0.04	-0.08	1.18	-0.07
Iteration 4	-0.04	0.38	-0.10	-0.15	0.69	-0.21	0.41	0.94	0.43	0.41	1.00	0.41
Iteration 5	-0.33	0.72	-0.46	-0.25	0.84	-0.30	-0.58	0.98	-0.60	0.50	0.79	0.63

*Note.* Pooled SD = pretest posttest pooled standard deviation; Cohen's *d* = measure of effect size.

<sup>a</sup>=Control group did not experience the Education research project.

Table 15

*Changes in Pretest and Posttest Scores on Items 6D, and 6E for Iterations 1-5*

	Research Articles (6D)			Practitioner Articles (6E)		
	$\Delta M$	Pooled <i>SD</i>	Cohen's <i>d</i>	$\Delta M$	Pooled <i>SD</i>	Cohen's <i>d</i>
Iteration 1A	0.06	0.91	0.07	0.42	0.97	0.43
Iteration 1B	0.42	0.87	0.48	0.42	0.93	0.45
Iteration 1C <sup>a</sup>	-0.29	0.82	-0.36	0.00	0.93	0.00
Iteration 2	0.37	1.04	0.35	0.63	1.15	0.55
Iteration 3	0.29	1.13	0.26	0.75	1.08	0.70
Iteration 4	0.15	0.77	0.19	0.19	0.90	0.20
Iteration 5	0.83	0.77	1.08	0.75	0.62	1.20

*Note.* Pooled *SD* = pretest posttest pooled standard deviation; Cohen's *d* = measure of effect size.

<sup>a</sup>Control group did not experience the Education research project.

**Inferential statistics.** In order to evaluate the change from pretest to posttest on items 6D and 6E for Iteration Five participants, I conducted two dependent-means *t*-tests. Prior to conducting the tests, I found no evidence of violation of the assumption of normality from inspection of values for skewness (ranged from -0.17 to -0.38) and kurtosis (ranged from 0.51 to 0.85). A paired samples *t* test indicated statistically reliable difference between the mean score of posttest ( $M = 4.08, SD = 0.67$ ) and the mean score of the pretest ( $M = 3.25, SD = 0.87$ ) for Item 6D,  $t(11) = 3.46, p < .025, \alpha = .025$ . A second paired samples *t* test indicated statistically reliable difference between the mean score of posttest ( $M = 3.92, SD = 0.51$ ) and the mean score of the pretest ( $M = 3.17, SD = 0.71$ ) for Item 6E,  $t(11) = 3.45, p < .05, \alpha = .05$ .

In order to compare the pretest-posttest change seen in Iteration 5 for items 6D and 6E with the pretest-posttest change for the first four iterations, I conducted two one-way ANOVAs on the difference scores for all five iterations. The first iteration included three separate sections, so the ANOVAs were run on seven course sections in total. Prior to conducting the ANOVAs, I found no evidence of lack of normality from skewness (6D = -0.07, 6E = 0.09) and kurtosis (6D = -0.23, 6E = 0.20) values and no evidence of violation of homogeneity of variance from the Levene's test (6D Levene's statistic = 0.67,  $p > .05$ ; 6E Levene's statistic = 0.57,  $p > .05$ ). Results of the one-way ANOVA for item 6D,  $F(6, 158) = 2.05, p > .05, \eta^2 = .07$ , failed to demonstrate that there were significant differences between the seven groups. Results of the one-way ANOVA for item 6E,  $F(6, 158) = 1.66, p > .05, \eta^2 = .06$ , also failed to demonstrate that there were significant differences between the seven groups. The eta squared effect sizes of .07 and .06 for item 6D and 6E, respectively, indicate medium effect sizes (Cohen, 1988).

I also compared change scores for Item 6D and 6E for participants in Iteration 1 and participants in Iteration 5 by conducting two independent-means *t*-tests. After testing for the relevant assumptions (skewness values ranged from -1.05 to 0.26; kurtosis values ranged from -0.51 to 2.06; 6D Levene's statistic = 0.16,  $p > .05$ , 6E Levene's statistic = 1.62,  $p > .05$ ), and using a modified Bonferroni approach to error rate control, I found a significant difference in means in favor of Iteration 5 on Item 6D,  $t(41) = 2.43$ ,  $p < .025$ ,  $\alpha = .025$ , but no difference in means on Item 6E,  $t(41) = 1.14$ ,  $p > .05$ ,  $\alpha = .05$ .

### **Familiarity with the Process of Evidence-Based Practice in Education Scale.**

Eleven out of 12 participants completed the pretest administration of the Familiarity with the Process of Evidence-Based Practice in Education Scale during the first week of the semester (Time 1). All 12 participants completed the posttest scale twice, at an interval of one week apart (Times 2 and 3), at the end of the semester. There were no missing item response data from these administrations. However, on the pretest one participant selected the response category "I do not know what this means" for item 4 and one participant selected that category for item 1. These two responses were recoded as "Neutral" or "3's" for the purpose of the quantitative analyses that follow. No participants selected the response category "I do not know what this means" at the posttest administration.

The final two administrations of the scale were used to calculate a test-retest reliability correlation coefficient of .79 for total scale score. Internal consistency was measured by calculating Cronbach's alpha. Due to the fact that sample size was quite low, I calculated the statistic for each of the three administrations in order to evaluate the consistency of the values; they were .79 ( $n = 11$ ), .83 ( $n=12$ ), and .80 ( $n=12$ ).

Table 16 provides descriptive statistics at the item level for pretest (Time 1) and posttest (Time 2) administrations of the scale. Total scores were calculated for each of the participants at pretest and posttest. After screening these data and finding evidence of only minor violations of the normality assumption (pretest skewness = -1.69, posttest skewness = -0.49; pretest kurtosis = 3.69, posttest kurtosis = -0.65; pretest Kolmogorov-Smirnov = 0.20,  $p > .05$ ; posttest Kolmogorov-Smirnov = .18,  $p > .05$ ) a dependent-means t-test was conducted and demonstrated that the posttest mean ( $M = 42.18$ ,  $SD = 3.89$ ) was statistically significantly higher than the pretest mean ( $M = 30.36$ ,  $SD = 5.32$ ),  $t(10) = 9.63$ ,  $p < .05$ .

In addition to evaluating change in the scale mean, I evaluated change in individual participant scores by calculating Reliable Change Index (RCI) scores (Jacobson et al., 2004). Figure 7 displays these scores in the form of a bar graph, and Figure 8 displays them as a box plot. The RCI scores ranged from 2.87 to 8.61 ( $M = 4.85$ ,  $SD = 1.67$ ).

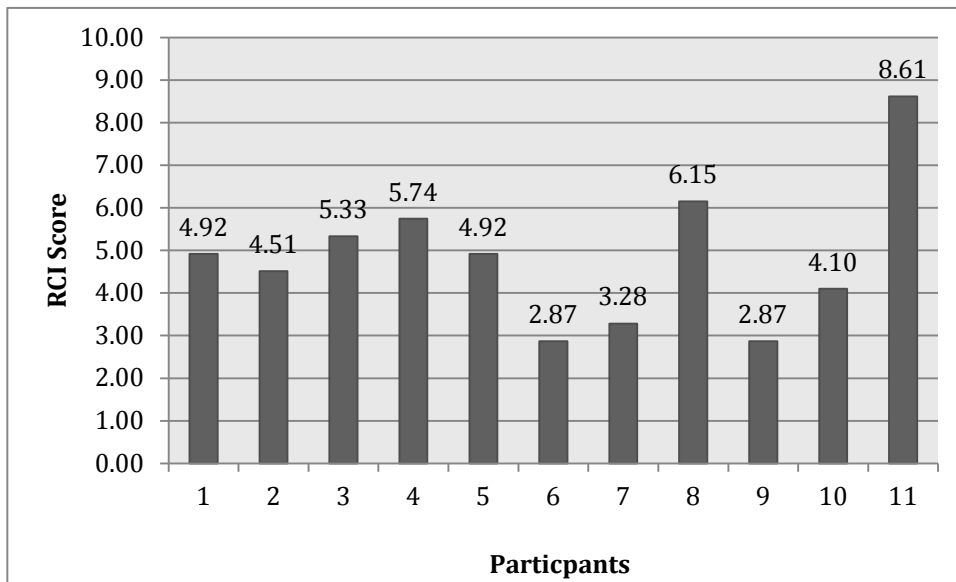
### **Education Research Project- Residency Version (ERP-RV)**

All 12 residents from Iteration 5 completed each of the 11 tasks for the ERP-RV. Table 17 provides a list of the topics that the residents chose to research for this project. A rubric (Appendix M) was used to score the projects. The descriptive statistics for the project scores are reported in Table 18. The ERP-RV (Table 10) required

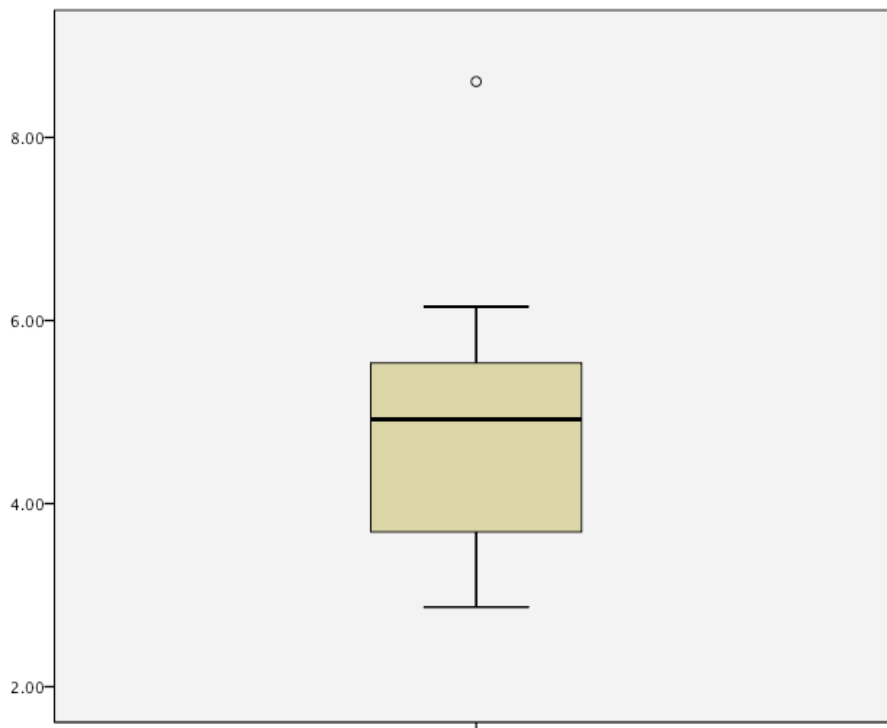
Table 16

*Item Level Descriptive Statistics for the Familiarity with the Process of Evidence-Based Practice in Education Scale*

Item	Pretest		Posttest	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
1. When I have a problem in classroom teaching, I can formulate a question that can be answered by education research.	3.73	0.91	4.42	0.52
2. Once I have a research question, I know how to look for research articles that address the question.	3.18	1.25	4.42	0.52
3. I can tell if an education article is theoretical or data-based.	3.18	1.25	4.50	0.67
4. I can tell if a data-based article uses quantitative, qualitative or mixed methodology.	3.00	1.18	3.83	0.84
5. I can judge whether a research article provides strong or weak evidence.	3.36	0.92	4.08	0.67
6. I can synthesize the evidence from several research articles.	3.18	0.98	4.58	0.52
7. I know how to create a plan to implement education research findings in my own classroom.	2.73	0.47	4.17	0.58
8. I am able to adapt the research findings to meet the needs of my particular students.	3.27	0.79	4.17	0.58
9. I know how to identify potential barriers that might hinder my attempt to implement education research in my classroom.	3.55	0.82	4.08	0.79
10. I know how to evaluate whether my use of research-based findings has been successful.	3.00	0.63	4.17	0.58



*Figure 7.* Graphical representation of reliable change index scores for eleven of the participants in Iteration 5.



*Figure 8.* Box plot of reliable change index scores for eleven of the participants in Iteration 5.



Table 17

*Research Topics and Number of Residents Researching the Topic for the Education Research Project- Residency Version*

Topic	<i>n</i>
Cooperative learning in mathematics class	1
Culturally responsive teaching in mathematics	1
Differentiation strategies to support students with low math achievement	1
Effective manipulative use	1
Effective use of technology in mathematics teaching	1
Fostering metacognition in mathematics	1
Integration of literature in mathematics class	1
Kinesthetic activity in mathematics class	1
The role of homework for mathematics learning	3
The role of parental involvement in mathematics homework	1

participants to create a written plan for how they intended to apply the results from their literature search to classroom teaching. This task of writing a research application essay was included in the previous versions of the ERP as well. Chapter 3 of this study describes how the research application essays from Iteration 1 were analyzed using grounded theory. The codebook that was created for this Iteration 1 analysis was used again to analyze the research application essays from Iteration 5. The use of the same coding scheme allows for comparison of the essays generated in these two different iterations.

After the research application essays from Iteration 5 were coded with the 40 codes, the essays from two iterations were compared across the four themes that developed from the grounded theory analysis of Iteration 1 (see Appendix E for the relationships between the codes and themes). These four themes were: Understanding of Research, Implementation Plan, Implementation Modification, and Implementation

Table 18

*Descriptive Statistics for Scores from the Education Research Project- Residency Version*

Rubric Item	Possible Points	<i>M</i>	<i>SD</i>
Was the research question related to the identified problem?	3	3.00	0.00
Was the research question manageable?	1	1.00	0.00
Did search terms cover all major and minor concepts in the research question?	3	2.92	0.29
Did student choose appropriate databases?	1	0.92	0.29
Did student choose appropriate articles?	3	3.00	0.00
Were articles cited in basic APA format (minor format variations ok)	3	2.58	0.51
Identified essential research findings for the 3 articles	3	2.75	0.62
Synthesis creates a coherent discussion of the similarities and differences among the articles.	3	2.75	0.45
Synthesis conclusion articulates essential points regarding the topic given the 3 articles	3	2.83	0.39
Written plan clearly connects intervention to stance on teaching and learning.	3	2.75	0.45
The implementation plan is clearly defined (actionable steps).	3	3.00	0.00
Plan is clearly connected to the research.	3	3.00	0.00
Plan gives consideration to local conditions: includes identification of possible barriers and supports to implementation.	3	2.17	0.72
Plan clearly articulates how the teacher will evaluate the implementation.	3	2.08	0.90
The reflection explains how the teaching was connected to the research.	3	2.83	0.39
Reflection explains how the teacher modified the research implementation to meet the needs of students.	3	2.92	0.29
Articulates effect of implementation on teacher and students.	3	2.67	0.49
Provides evidence of success/lack of success of implementation based on video or artifacts.	3	2.33	0.49
Makes reasonable recommendations for future implementation.	3	3.00	0.00
Total Score	53	48.50	3.26

Evaluation. Table 19 shows how the original codes were collapsed to form dichotomous success and lack of success categories for each of the four themes: understanding of research, implementation plan, modification of implementation, and evaluation of implementation. Figure 9 provides graphical comparisons of the rates of success for Iteration 1 and Iteration 5 on each of the four categories. There is a consistent pattern of greater success for Iteration 5 participants than Iteration 1 participants on each of the four categories.

Table 19

*Relationships of Original Codes to Dichotomous Success Categories for Four Themes from the Research Application Essays of Iterations One and Five*

Theme	Codes Indicating Success	Codes Indicating Lack of Success
Understanding of Research	<ul style="list-style-type: none"> <li>• Identifies essential research findings</li> </ul>	<ul style="list-style-type: none"> <li>• Does not identify essential research findings</li> </ul>
Implementation Plan	<ul style="list-style-type: none"> <li>• Clearly defined implementation plan</li> </ul>	<ul style="list-style-type: none"> <li>• Does not identify implementation plan</li> <li>• Global goal but no implementation plan</li> <li>• No implementation plan</li> <li>• Poorly defined implementation plan</li> </ul>
Modification of Implementation Plan	<ul style="list-style-type: none"> <li>• Considers local conditions</li> <li>• Identifies implementation barrier</li> <li>• Identifies potential support to implementation</li> <li>• Recognizes need for iteration of implementation</li> </ul>	<p>(The absence of success codes is considered lack of success for this theme)</p>
Evaluation of Implementation Plan	<ul style="list-style-type: none"> <li>• Has a plan to evaluate implementation</li> </ul>	<ul style="list-style-type: none"> <li>• No plan to evaluate implementation</li> </ul>

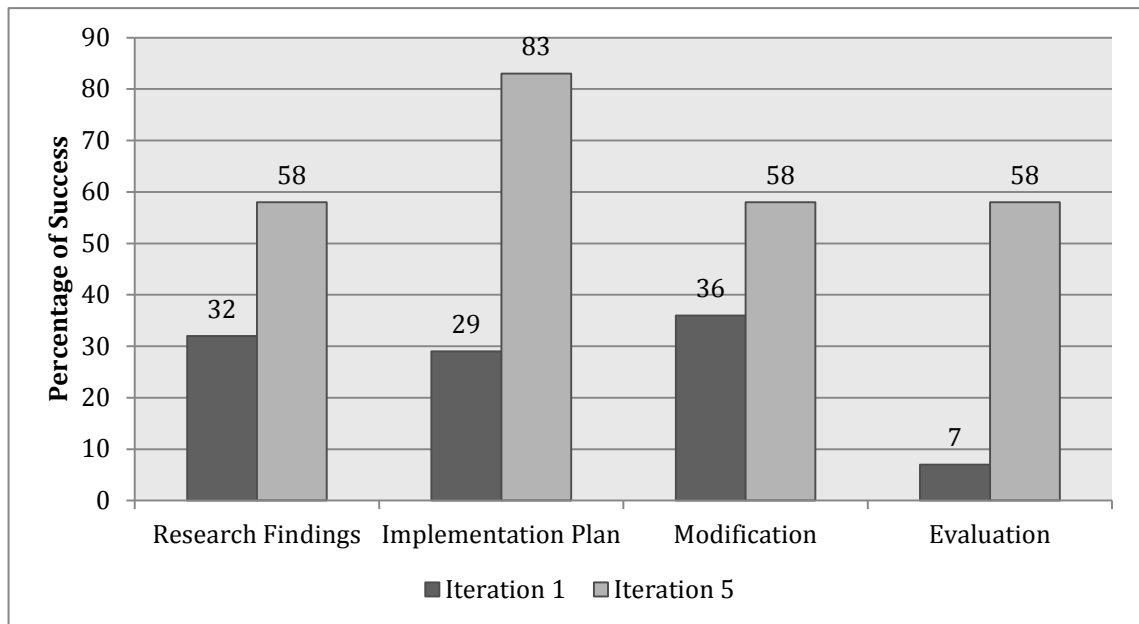


Figure 9. A comparison of Iteration 1 and Iteration 5 participant success rates on communication of four categories related to the process of applying research to practice in the research application essays from the Education Research Project. Note that for Iteration 1,  $n = 28$  and for Iteration 5,  $n = 12$ .

### Standardized Performance Assessment

The extent to which the standardized performance assessment was sensitive to instruction was evaluated through a dependent means  $t$ -test of the average pretest and posttest performance assessment essay scores. After screening these data and finding no evidence of violation of the normality assumption (pretest skewness = 0.29, kurtosis = -0.57, Shapiro Wilk = .95,  $p > .05$ ; posttest skewness = -0.17, kurtosis = -0.42, Shapiro Wilk = 0.97,  $p > .05$ ), a dependent-means  $t$ -test was conducted and demonstrated that the posttest mean ( $M = 14.10$ ,  $SD = 2.24$ ) was statistically significantly higher than the pretest mean ( $M = 9.92$ ,  $SD = 1.30$ ),  $t(11) = 5.92$ ,  $p < .05$ . Although the total score on the standardized performance assessment is of greatest interest, Table 20 displays the average, item-level scores for this assessment.

Table 20

*Descriptive Statistics for Item-Level Scores from the Standardized Performance**Assessment*

Rubric Item	Possible Points	<i>M</i>	<i>SD</i>
The implementation plan is clearly defined (actionable steps).	3	2.52	0.51
The plan is clearly connected to the research.	3	2.83	0.21
The plan articulates the need to make modifications to research implementation to meet the needs of the particular students in the classroom.	3	1.80	0.53
The plan anticipates a barrier to the implementation of the research findings.	3	1.73	0.46
The plan provides a potential remedy/support to anticipated barriers to implementation.	3	1.47	0.46
Plan clearly defines how the teacher will evaluate the implementation.	3	1.40	0.62
The plan for evaluation of the implementation could reasonably be carried out by a classroom teacher.	3	1.44	0.67

*Note.* Means are grand means computed from item means of 12 participants. Each participant's mean computed from 12 scores: 2 tasks x 2 scoring occasions x 3 raters.

Generalizability theory was used to evaluate the sources of error and estimate the reliability of the system created for the standardized performance assessment. Table 21 reports the estimated variance components and percent of variance for each of the sources of variance in the fully-crossed three facet generalizability study design. The variance component estimates were used to calculate a relative error variance of .63 and a relative generalizability coefficient of .85. The absolute error variance was .85 and the absolute generalizability coefficient was .81. The generalizability coefficients are analogous to reliability coefficients and provide a sense of score dependability.

Table 21

*Sources of Variance in the Three-Facet Generalizability Study*

Source of Variance	<i>n</i>	Estimated Variance Component	Percent Variance
Persons (p)	12	3.526	51.59
Rater (r)	3	0.473	6.92
Task (t)	2	0.000	0.00
Occasion (o)	2	0.011	0.16
p x r		0.950	13.90
p x t		0.192	2.81
p x o		0.000	0.00
r x t		0.130	1.90
r x o		0.000	0.00
t x o		0.000	0.00
p x r x t		0.960	14.05
p x r x o		0.000	0.00
p x t x o		0.041	0.60
r x t x o		0.000	0.00
p x r x t x o, e		0.551	8.06
Total		6.834	100.00

Following the generalizability study, I conducted a series of decision studies to understand how manipulation of the testing conditions would alter the relative and absolute generalizability coefficients. The results of the decision studies are presented in Table 22. This table quantifies the consequences to score dependability when testing conditions are altered. Holding task and scoring occasions both at two, one can see that the maximum reliability (absolute G-coefficient) for two raters is only .75 and increases to .81 with three raters and .85 with four raters. This table can be used for a cost-benefit

analysis of the expense of testing conditions and the desired score dependability.

Although space considerations limit the size of the table, it is worth noting that the common assessment conditions of a single rater (teacher educator) and a single scoring occasion would require 21 tasks (essays) in order to reach a reliability coefficient (absolute G-coefficient) of .70. Even the impossibly high condition of 1000 essays with one rater and one scoring occasion would not be sufficient to raise the G-coefficient to the more acceptable level of .80.

Table 22

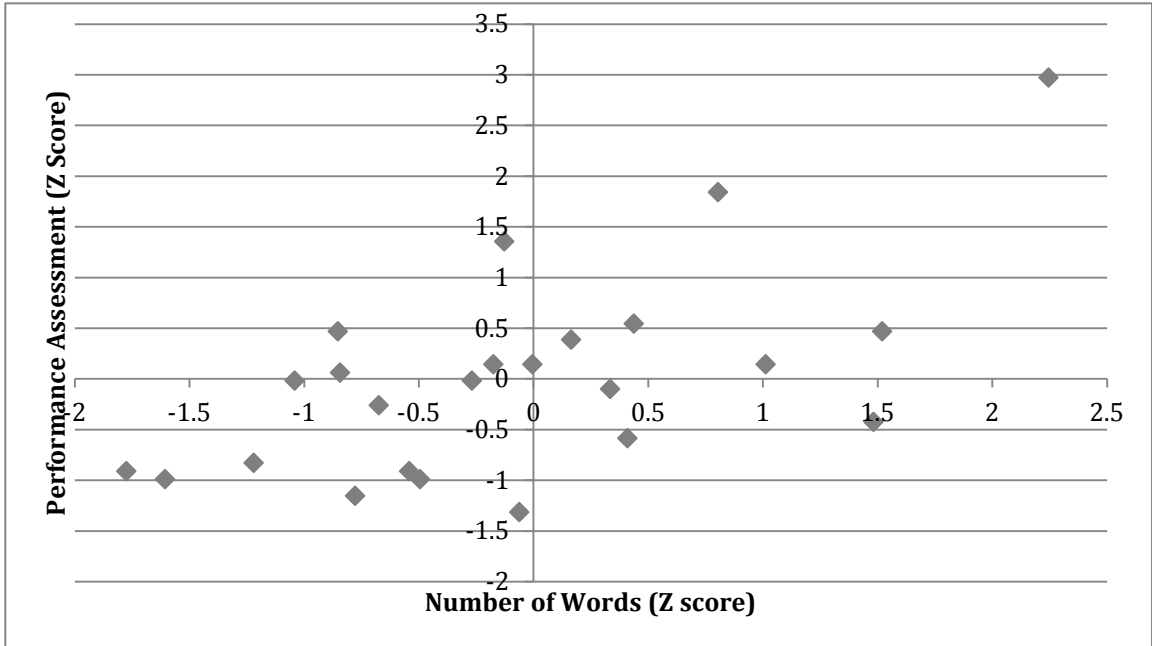
*Relative and Absolute Generalizability Coefficients from Decision Studies in which the Testing Conditions of Raters, Tasks, and Occasions were Manipulated.*

Raters	Tasks	Occasions	Relative Error Variance	Relative G-Coefficient	Absolute Error Variance	Absolute G-Coefficient
4	2	2	0.498	.88	0.638	.85
4	1	2	0.759	.82	0.915	.80
4	1	1	0.848	.81	1.010	.78
3	2	2	0.629	.85	0.814	.81
3	1	2	0.941	.79	1.148	.76
3	1	1	1.053	.77	1.265	.74
2	2	2	0.890	.80	1.165	.75
2	1	2	1.305	.73	1.612	.69
2	1	1	1.464	.71	1.776	.67
1	2	2	1.674	.68	2.218	.62
1	1	2	2.398	.60	3.007	.54
1	1	1	2.694	.57	3.308	.52

To explore the extent to which writing ability may have influenced the performance assessment scores, I invited the residents to participate in brief interviews to discuss their responses to one of the performance assessment essays. Only one resident volunteered to participate in the interview. The protocol used for the interview can be found in Appendix P. On the standardized performance assessment, the resident was unsuccessful in identifying modifications to the implementation plan and was also unable to communicate a plan for modification after an initial verbal prompt. However, upon further prompting, she was able generate an idea for implementation modification. Likewise, with the initial prompt, the resident was not able to provide a plan to evaluate the implementation. However, after additional prompts the resident had an “aha!” moment and generated an effective plan for evaluation of the implementation.

Due to the low volunteer rate for the follow-up interviews, I looked for alternative sources of information on the effect of writing ability on the standardized performance assessment. I examined how strongly the word count was correlated to the score of the performance assessment. For each resident, I calculated the number of words per essay and the average score from the three raters for each of the two standardized performance assessment essays. The Pearson product moment correlation coefficient between word count and assessment score was .61. This means that, in the context of considering only word count and assessment score, approximately 38% of the variance in the assessment score can be attributed to essay word length. Figure 13 shows a scatterplot of the relationship between word length and performance assessment score.





*Figure 10.* The relationship between the number of words and the score for Iteration Five standardized performance assessments.

## **Chapter 5: Discussion**

The purpose of this chapter is to discuss both the interpretations and implications of the results presented in Chapter 4. To accomplish this purpose, I have organized the discussion into five sections. First, I revisit the research questions from Chapter 1 and discuss the extent to which the results from Iteration 5 answered these questions. Second, I revisit the literature review from Chapter 2 and discuss the research implications of the DBR study in the context of the literature base. Third, I discuss the practice implications from the study. Fourth, I provide directions for future research. Finally, I offer concluding statements.

### **Interpretation of Results**

There were four research questions that guided the design of Iteration 5. These research questions were:

- 1) After experiencing the intervention, to what extent did the participating preservice teachers demonstrate the ability to apply research to their classroom teaching?
- 2) To what extent did preservice teachers' intentions to apply research to future teaching change from the beginning to the end of the intervention?
- 3) What, if any, were the differences in a) ability to apply research to practice and b) intention to apply research to practice between residents in the fifth iteration and participants in previous iterations?
- 4) How reliable was the system of assessment that was used to assess preservice teachers' abilities to apply research to teaching?

I will address each of these questions in order and will interpret the relevant results from Chapter 4 in terms of the extent to which they answer the questions.

**Question 1.** There were three sources of data that were designed to address the extent to which Iteration 5 residents demonstrated the ability to apply research to classroom teaching: (a) the products from the Education Research Project-Residency Version (ERP-RV), (b) the standardized performance assessment, and (c) the Familiarity with EBP in Education Scale. I now interpret in turn the results from each of these data sources.

Table 18 in Chapter 4 documents that the residents were highly successful in the completion of the ERP-RV. The average total score on this project was 92%, and, of the 19 criteria used for evaluation purposes, all 12 of the residents received the maximum score possible on the following six criteria: research question was related to a classroom problem, the research question was manageable, appropriate articles were chosen, the implementation plan was clearly defined, the plan was clearly connected to the research, and the resident made reasonable recommendations for future iterations of implementation. However, there were three criteria for which the average score was below 2.5 out of 3.0. These were: the plan gives consideration to local conditions, the plan clearly articulates how the teacher will evaluate the implementation, and the resident provided evidence of success of implementation.

The overall success of residents on the ERP-RV can be interpreted as proof-of-concept that, given the support conditions present in this study, this group of residents was capable of engaging in the process of applying research to teaching. Using cross-disciplinary terminology, residents were capable of engaging in the process of evidence-

based practice (EBP). The importance of the supportive context of the ERP-RV must be emphasized in this interpretation. The residents were coached by both an education research librarian and a course instructor throughout the process of completing the ERP-RV. Furthermore, an extended time period, an entire semester, was provided to the residents for the completion of this project.

Although the study conditions supported a high level of resident success on most of the ERP-RV, it is also important to note that the conditions were not sufficient to support the same level of success for resident ability to modify research plans to meet local conditions (specific classroom needs) and to plan for and carry out evaluation of the implementation plan. These lower rates of success correspond to steps four and five of the process of EBP (see Table 4) and ought to be considered as a potential area of revision for future use of the ERP-RV.

Whereas the ERP-RV provided information about residents' abilities to engage in the five steps of EBP under the conditions of coaching and an extended time period, the performance assessment provided information about residents' abilities to engage in EBP steps four and five without assistance and within a brief time period. Furthermore, this assessment was a standardized approximation of steps four and five. All residents read the same research summary and were required to create written plans for implementation and evaluation of that research.

Because the instructional sensitivity comparison of pre post administrations of the performance assessment showed statistically significant growth in post assessment scores, it is plausible that experience with the ERP-RV intervention had a positive impact upon resident ability to succeed on the performance assessment. Nevertheless, additional

research (an experimental study with random assignment to a control group) would be necessary to make this conclusion. Although causal statements cannot be made as to what caused the level of performance on the posttest standardized performance assessment, it is appropriate to look at the descriptive statistics of that performance as evidence of the posttest capabilities of the residents.

Table 20 indicates that the residents were more successful in defining an implementation plan and connecting that plan to research than they were in anticipating necessary modifications to their implementation plans and evaluating their plans. This is consistent with the pattern of success rates seen in the ERP-RV. Table 23 allows for a side-by-side comparison on four similar tasks that were rated on both the ERP-RV and the standardized performance assessment. The Pearson product moment correlation coefficient between the residents' scores on the four ERP-RV items and the corresponding four SPA items was .48. This was a moderate correlation and represented a non-trivial degree of association between the two sets of scores.

Caution must be exercised in the direct comparison of mean scores due to the fact that different scoring criteria were employed for each tool (see Appendices J and M for the scoring rubrics). Despite the need for caution in direct, quantitative comparisons of the means, the scoring patterns can be analyzed. In addition to the fact that residents scored lower across both tasks on modification and evaluation of the implementation plans, residents also scored consistently lower on the performance assessment than on the ERP-RV. This pattern of lower performance on the standardized performance assessment is significant because it possibly indicates a gap between supported, time-

extended resident engagement in EBP (ERP-RV) and unsupported, time-limited engagement in EBP (performance assessment).

Whereas the ERP-RV and the standardized performance assessment provide documentation regarding resident performance on tasks that approximate the process of EBP, the Familiarity with Evidence-Based Practice in Education Scale provides self-report data about how residents perceive their ability to engage in EBP. Prior to the

Table 23

*Four Points of Comparison Between Average Scores on the Education Research Project-Residency Version and Scores on the Standardized Performance Assessment*

		Rubric Item	Possible Points	M	SD
1	ERP-RV	The implementation plan is clearly defined (actionable steps).	3	3.00	0.00
	SPA	The implementation plan is clearly defined (actionable steps).	3	2.52	0.51
2	ERP-RV	Plan is clearly connected to the research.	3	3.00	0.00
	SPA	The plan is clearly connected to the research.	3	2.83	0.21
3	ERP-RV	Plan gives consideration to local conditions: includes identification of possible barriers and supports to implementation.	3	2.17	0.72
	SPA	The plan articulates the need to make modifications to research implementation to meet the needs of the particular students in the classroom.	3	1.80	0.53
4	ERP-RV	Plan clearly articulates how the teacher will evaluate the implementation.	3	2.08	0.90
	SPA	Plan clearly defines how the teacher will evaluate the implementation.	3	1.40	0.62

*Note.* ERP-RV = Education Research Project- Residency Version; SPA = Standardized Performance Assessment. Caution must be exercised in comparison of ERP-RV and SPA scores as different scoring criteria were used in each scoring rubric.

interpretation of scores, it is important to consider the reliability of the scores. Although the test-retest reliability coefficient for total scale score (.79) represents an acceptable level of test-retest reliability, it must be interpreted with caution due to the small sample size of participants ( $n=12$ ).

The statistically significant positive change in total scale score from pretest to posttest indicates that the residents perceived themselves as more capable of engaging in the process of evidence-based practice at the end of the semester than at the beginning of the semester. This growth is congruent with the growth seen in the pretest posttest performance assessment scores. However, at posttest residents provided the same mean level of endorsement for the skills of creating an implementation plan, modifying the plan, and evaluating the plan. This is in contrast to the results of the ERP-RV and performance assessment that indicated less capacity to modify and evaluate a plan as compared to the capacity to create an implementation plan. The incongruity between the self-report data and the performance data suggests the possibility that residents might benefit from additional feedback or information regarding their demonstrated ability to modify and evaluate implementation plans.

The reliable change index (RCI) scores reported in Figure 7 illustrate the fact that the growth in self-reported capacity to engage in the process of EBP as measured by the Familiarity in Evidence-Based Practice in Education Scale was not uniform but varied across participants. The highest RCI score (8.61) was about three times larger than the lowest two scores (2.87). Unfortunately, the number of participants is too small to establish a cut off point that indicates reliable change, but even without the cut off score, it is clear that participants experienced quite different rates of change. Additional

research is needed to understand why and how some residents experienced more change than other residents.

In conclusion, there were three sources of information that addressed the first research question. All three sources of information indicated that the residents demonstrated a capacity to engage in the process of EBP. The ERP-RV provided proof-of-concept evidence that, with support and extended time, residents were able to demonstrate approximations of each of the five steps of EBP. Both the standardized performance assessment and the Familiarity with Evidence-Based Practice in Education Scale provided evidence of greater capacity at the end of the semester as compared to the beginning of the semester. The performance data (ERP-RV and standardized performance assessment) indicated that residents were relatively less able to modify and evaluate their implementation plans as compared with their capacity to create implementation plans. In contrast, the self-report data indicated that the residents saw themselves as equally capable of creating, modifying, and evaluating implementation plans.

**Question 2.** Residents' intentions to apply research to practice was assessed through Items 6D and 6E on the Information Literacy Questionnaire. These items asked the residents to indicate, on a scale of 1 to 5, the likelihood that they would use research articles and practitioner articles, respectively, to inform classroom teaching. The item-level test-retest reliability coefficients for these items were low at .45 and .62. Again, these coefficients must be interpreted with caution due to the fact that sample size was very low ( $n=11$ ). In addition, the low reliability coefficients indicate caution should be exercised in the interpretation of pretest posttest change scores for these items. In Table



15, for item 6D it is reported that the change in mean from pretest to posttest was 0.83. This corresponds to a very large Cohen's *d* effect size of 1.08. For item 6E, the pretest posttest change was 0.75 with a Cohen's *d* value of 1.20. These large effect sizes suggest that there was a meaningful increase in endorsement of the use of research to inform teaching from the beginning of the semester to the end of the semester.

Remembering that the low item reliability levels indicate caution should be exercised in interpretation of these large effect sizes, it is also possible to compare the effect sizes for changes on items 6D and 6E, changes which were anticipated to be strong and positive due to the EBP intervention, with changes in the four items that asked about other resources that teachers might use to inform teaching practice. From Table 14, it is clear that the changes in endorsement of using advice from other teachers, advice from administrators, information from Google searches, and information from course textbooks (-.46, -.30, -.60, and .63 respectively) were not as positive nor as strong as the changes for items 6D and 6E. These resources were not specifically targeted during the semester-long intervention. Thus, this evidence contributes to a plausible interpretation that the experiences of the semester surrounding engagement of the process of EBP were associated with an increase in residents' intentions to use research to inform their teaching practices.

The above interpretations of the findings from the Information Literacy Questionnaire were limited due to low item-level test-retest reliability values. Future use of this questionnaire could be strengthened if the number of items that assess teacher intentions to use specific resources to inform practice were increased. Increasing the number of high quality assessment items could lead to greater score reliability.

**Question 3.** The design of Iteration 5 of the DBR agenda was significantly different from the design of previous iterations in that the participants were embedded in K-5 classrooms while enrolled in the mathematics methods course. The first nine tasks of this education research project (ERP-RV) were similar to tasks in previous iterations. Yet, the ERP-RV went beyond prior versions because it included additional requirements for the residents to implement their research-based plans in classroom teaching and to evaluate the implementations. Did the simultaneous classroom placement and the opportunity to implement the research plan impact the quality of the written implementation plans? Was the different environmental context associated with greater intention to apply research to practice? In order to answer these questions, I compared responses from Iteration 5 participants to responses from Iteration 1 participants on the written implementation plans (a similar task on both versions of the education research projects) and pre-post changes in items 6D and 6E on the information literacy questionnaires.

The analysis of the written implementation plans revealed a consistent pattern of greater success in approximating each of steps three, four, and five in the process of EBP for participants in Iteration 5 than participants in Iteration 1 (See Table 4 for the five EBP steps). Figure 9 provides a visual display of this pattern. Step three of the process of EBP requires the teacher to appraise critically the research evidence. Detailed coding of the implementation essays revealed that approximately one third of Iteration 1 participants communicated the essential research findings that informed their research implementation plans. In contrast, more than half of participants in Iteration 5 were successful in communicating these findings. Similarly, approximating step four of the

EBP process, less than one third of Iteration 1 participants demonstrated the ability to define clearly a research implementation plan whereas more than four fifths of the Iteration 5 participants did so. In addition, Iteration 1 participants were less likely to indicate how they might modify their plans to meet the needs of their specific students. Less than one tenth of Iteration 1 participants communicated such an evaluation plan as compared to more than fifty percent of Iteration 5 participants. In summary, on each of these four points of comparison that are approximations of steps in the process of EBP, participants in Iteration 5 demonstrated greater proficiency than participants in Iteration 1. Although such a finding does not *prove* that the intervention in Iteration 5 was superior to that in Iteration 1, it is consistent with such a conclusion.

The second part of research question three involved an inquiry into any difference between Iteration 1 and Iteration 5 on change in self-reported likelihood to use research to inform future teaching practice (Items 6D and 6E on the Information Literacy Questionnaire). As was mentioned in the previous discussion for research question two, the results of this analysis must be interpreted with caution as the test-retest reliability coefficients for these two items were low. With this caveat in mind, I found that participants in Iteration 5 indicated a greater change in likelihood to use scholarly research articles to inform teaching practice than those participants in Iteration 1 (Item 6D) but there was no difference in change on likelihood to use practitioner articles (Item 6E). Potentially, this finding could indicate that the intervention in Iteration 5 was more potent than that of Iteration 1. If this finding were replicated in future research with scores that demonstrated more reliability, it would also be useful to alter the research design so that one could parse out if this increase in potency was due to the fact that

participants in Iteration 5 had more exposure to education research in general or to the fact that they had the opportunity to apply the research to practice, or a combination of these two factors.

**Question 4.** The generalizability study demonstrated that the system of assessment established for the standardized performance assessment produced scores that were fairly reliable. The relative generalizability coefficient of .85 can be interpreted as the reliability of correctly ranking participants' performance assessment scores.

Although Hill et al. (2012) suggested that the relative coefficient may be of primary interest in the context of merit pay tied to school district performance assessments, in this context of learning foundational teaching practices in a teacher preparation program, the absolute generalizability coefficient is of greater interest. In the context of teacher preparation, the teacher educator is most often concerned about proficiency in performance versus the ranking of performances. It is the absolute generalizability coefficient that identifies the reliability of the scores themselves (versus the ranking of the scores). In this study, the absolute reliability coefficient of .81 indicated a fairly high level of performance assessment score reliability.

A great advantage to conducting a generalizability study is that, in addition to finding an overall reliability coefficient, one is able to identify the various sources of error in the score variance (see Table 21). Of the three facets included in this study (rater, task, and scoring occasion), the facet of rater contributed to the greatest source of error. In other words, the facet of rater is the optimal facet to target if one would like to increase the reliability of the scores. In addition to reducing error through the use of

additional raters, the error related to rater might be reduced through improved rater training.

Scoring occasion was the second largest source of score error. Each rater scored each performance twice. Thus, scoring occasion can be interpreted as a measure of intra-rater reliability. Again, in addition to having raters score performances on more than two scoring occasions, this source of error might also be reduced through additional rater training. Furthermore, the variance due to scoring occasion might be an indication that the rubric could be refined further such that the scoring categories are even clearer and the cognitive load is reduced for the scorer. This might enhance the intra-rater reliability (Jonsson & Svingby, 2007).

Perhaps the most surprising finding in this study is that the main effect of task did not contribute as a source of error variance. Task refers to the specific research summary about which residents read and created their written research implementation plans. In this study, there were two different tasks that the residents completed as part of the standardized performance assessment. One might have predicted that, due to differences in residents' prior knowledge, performance would have varied based upon resident familiarity with the specific research topic. Indeed, studies across a range of disciplines have found that task selection accounts for a significant source of error variance (e.g. Boulet, 2003; Brennan, 2000; Shavelson et al., 1991). Nevertheless, there are two plausible reasons why the main effect of task did not contribute to error variance. First, both research summaries that were chosen addressed foundational topics to elementary mathematics teaching. One topic related to the teaching of fractions and another topic related to cooperative learning in mathematics. Second, and perhaps of most relevance,

the trichotomously scored rubric was designed only to assess whether or not the resident's performance included essential elements of steps four and five of the process of evidence-based practice. Specifically, the rubric identified if the residents created a research plan, if they included plan modifications to meet the needs of specific students, and if they included a means of evaluating the research implementation. For the most part, the rubric did not analyze the quality of these steps. There is precedence in the literature for simple rubrics being associated with stability of the task facet (Conigliaro & Stratton, 2010).

Although the system of conditions established in this study was sufficient for supporting an acceptable level of score reliability, I conducted a series of decision studies in order to understand better how manipulation of the facet levels might influence score reliability. Table 22 displays the results of these studies. Score reliability is costly, in terms of both financial and human resources. The findings on Table 22 allow the user of this performance assessment to make informed decisions regarding the impact of altering facet levels due to resource constraints. Due to the fact that these decision studies were carried out in the context of a DBR study in which feasibility and usability are highly valued, I used maximum values of four for rater, two for task and two for occasions. In the context of the teacher preparation program in which this study took place, it did not seem feasible to have more than four raters, to ask those raters to rate each task more than twice, or to ask residents to complete more than two tasks for a performance assessment. Both the time of the mathematics educators and the time of the residents in this program were extremely limited. Table 22 allows future users of the standardized performance

assessment to understand the relationships between cost and score reliability and to make an informed cost-benefit decision regarding testing conditions.

Finally, in addition to collecting and analyzing data on the standardized performance assessment score reliability, I also collected some data on one potential threat to score validity—the quality of writing. Because only one resident volunteered for a follow-up interview, it was not possible to understand the extent to which residents' written expression matched their verbal expression. As a proxy for quality of written expression, I also examined the extent to which word count was related to performance assessment score. The correlations between word count and score was .61. This is a substantial correlation and signifies that, in general, those who wrote more scored higher. Nevertheless because word count is only a proxy for writing quality, it is still difficult to understand the extent to which writing quality may or may not have been a threat to score validity. In addition, when evaluating this potential threat to validity it is important to consider the design of the rubric, which was created to evaluate a beginning level of proficiency in EBP. The rubric was designed to provide greater penalties for omission of EBP steps than for the inclusion of ineffective or poorly designed steps. Further validation studies might collect interview data or classroom observations of EBP implementation as means of evaluating the potential validity threat that quality of written expression poses to the scores of the standardized performance assessment.

### **Research Implications**

In the previous section, I interpreted the results of the Iteration 5 data analyses in light of the four research questions that guided the design of the iteration. Having answered these questions, I now situate the study findings in terms of the larger literature

base. I discuss the research implications of the findings for each of the five areas of literature that were reviewed in Chapter 2: preparing practitioners for evidence-based practice, linking research to practice in teacher preparation, routines of practice in mathematics teacher preparation, performance assessments, and design-based research. These research implications are designed to address the interests of education researchers.

**Preparing practitioners for evidence-based practice.** This study contributes to a cumulative body of research on the process of EBP because it adapted a well-established conceptual framework that has been used in the context of social work and other health related disciplines and applied the framework to a new context of teacher preparation. The intervention developed during the study, particularly the Education Research Project, provides EBP researchers with a specific, documented model of how the EBP steps can be approximated and decomposed for novice practitioners. Furthermore, the assessment data, both the Familiarity with Evidence-Based Practice in Education Scale and the Standardized Performance Assessment, documented the learning gains that might be expected from those who experience the intervention.

Consistent with findings from meta-analyses (Coomarasamy & Khan 2003; Khan & Coomarasamy, 2006), this study found that clinical integration of EBP training (Iteration 5) resulted in superior learning outcomes to coursework-based training (Iteration 1). Furthermore, in producing a system of assessment with detailed score reliability evidence, this study responded to a call for carefully developed EBP assessments (Flores-Mateo & Argimon, 2007).

This study also identified the final two steps of the EBP process as particularly challenging for novice practitioners. Iteration five participants were less effective both at



communicating plans to modify research implementation to meet the needs of specific students and communicating plans to evaluate research implementation. In the context of social work, Adams et al. (2009) identified a difficulty similar to that of modifying research plans when they discussed the challenge for social workers of balancing clinical wisdom with research findings. They suggested that there were no research-based guidelines for such a balance and that additional research was needed. This study would contribute to the conclusion that additional research is needed in this area.

The finding that Iteration 5 participants' self-report data did not reflect a lack of skill in implementation modification or evaluation indicated that the participants may have needed more feedback about their EBP performance. Existing literature in EBP suggests that providing fieldwork-based feedback for those learning the EBP process is essential but difficult to achieve (Franklin, 2007; Mullen, Bellamy, Bledsoe, & Francios, 2007; Proctor, 2007; Rubin, 2007).

**Linking research to practice in teacher preparation.** This study contributes to the literature on linking research to practice in teacher preparation by documenting the complexity that exists in the process of linking research to practice. It has been common for education researchers to evaluate teacher attitude toward education research and its applicability to practice (e.g., DeGeest, 2010; Gitlin et al., 1999; Ratcliffe et al., 2005). This study highlights the fact that a teacher's attitude toward research is distinct from the teacher's ability to engage in the process of applying research to practice and that the ability to apply research to practice is composed of several distinct skills. Participants in Iteration 5 of this study strongly endorsed a willingness to use education research to inform teaching practice, and yet the performance assessment of the skills necessary for

this process indicated that these participants still needed to develop skills related to modifying and evaluating research implementation.

It has been discussed that teachers resist the mandated use of practices supported by research (Lamb et al., 2007) and that teachers are unable to draw implications for teaching from research reports (Schmidt, 2011). In light of the current study, these findings are understandable and predictable. It is unsurprising that teachers are unable to engage in a complex task for which they have been unprepared. For the purposes of moving the field forward, it would be more productive for educational researchers to focus on the equipping of teachers to link research to practice rather than the documenting of a lack of ability to do so.

Numerous studies have documented efforts to teach preservice and inservice teachers how to engage in specific practices that are supported by research (e.g., Allinder, 2001; Bain et al., 2009; Klinger et al., 2003). This study offers an alternative research agenda that might be more fruitful and perhaps more potent in bridging the research to practice gap. Instead of devoting resources to researching effective training conditions for specific research-supported strategies, education researchers might consider devoting resources to studying the generic, meta-process of applying research to practice. This alternate research agenda has the potential for more widespread impact.

Everett et al. (2008), Harrison et al. (2006), and Sinnema et al. (2011) are three examples of studies that have specifically engaged in preparing teachers or preservice teachers to apply research to practice. None of these studies adequately described the preparation process in such a way that it could be replicated by others. In addition, each of these studies lacked a theoretical or conceptual foundation for the design of the

intervention to help link research to practice. The current study advances the field forward by explicating a specific conceptual framework for the process of EBP and providing detailed documentation about how that framework can be put into practice.

**Routines of practice.** The findings from this study contribute to the emerging body of literature that aims to develop a pedagogy of practice for teacher preparation (Grossman et al., 2009). Presented in Chapter 2, Table 3 provided a theoretical argument for why the process of EBP fulfills Ball et al.'s (2009) criteria for a high leverage practice in the field of mathematics education. In addition to this theoretical argument, the findings from this study provided empirical evidence that the process of EBP does indeed meet the definition of a high-leverage routine of practice. Specifically, the findings demonstrated that the process of EBP can be articulated and taught, is accessible to the learners of teaching, can be revisited at increasingly sophisticated levels, and can be practiced in field-based settings.

In addition to the identification of a specific new routine of practice, this study contributes to the knowledge base about how, in general, a routine of practice can be decomposed, approximated, and assessed in the context of a teacher preparation program. The five-step conceptual framework provided an initial decomposition of the complex task of linking research to practice, but the study intervention further decomposed these five steps into a series of eleven specific tasks (see Table 10 for a visual representation of the relationships between the steps and tasks).

The documented development of the intervention from iteration to iteration exposed the process in which teacher educators engage when identifying the most effective approximations of a particular routine of practice. For example, the intervention

in this study progressed from being composed primarily of the ERP to involving a micro-teaching application of the ERP to involving a field-based application of the ERP. The decision to refine the approximation of practice was based upon an evaluation (in this case the treatment potency proxy test) of the effectiveness of the intervention.

As Chapter 2 noted, there has been little literature on the assessment of routines of practice. Thus, this study contributes to the literature base by providing detailed documentation of a system of assessment and the reliability of its scores for a routine of practice. The design of the rubric used for this performance assessment reflects Moss's (2011) recommendation to attend to conceptions of quality. In fact, the rubric is based upon a low expectation of quality that is appropriate for assessing the beginning stages of learning a routine of practice. In general, the rubric assesses that the novice is *aware* of and can communicate steps critical to the process of EBP. As teacher preparation programs focus more on developing proficiency in this routine of practice, additional assessments that measure greater levels of proficiency in this routine would need to be developed. This study's explicit attention to the creation of a system of assessment and the use of generalizability theory, as recommended by Hill et al. (2012), provide an example for educational researchers of how future performance assessments for routines of practice might be developed.

**Performance assessments.** Because this DBR study produced not only a teaching intervention, but also a system of assessment to support that intervention, the study contributes, in a small way, to the literature on performance assessments. The reporting of the assessment conditions and the use of generalizability theory add to the literature on the factors that influence score reliability. A prior meta-analysis (Hoyt &

Kerns, 1999) has shown that the facet of rater contributed to 37% of score variance. In this study, the facet of rater alone contributed to about 7% of the variance and interaction terms that included rater accounted for about 38% of the variance. As discussed in the prior section on research interpretations, this study found that the task selection main effect was not a significant source of score variation. This is in contrast to the findings of studies such as Boulet (2003), Brennan (2000) and Shavelson et al. (1991) but similar to the findings of Conigliaro and Stratton (2010). One hypothesis for this finding is that the general nature of the rubric accounted for little score variation attributed to task specificity. In this study the source of true score variance, the preservice teacher, accounted for 52% of the variance. This is a higher percentage than has been reported for other recent teacher performance assessments (Hill et al., 2012; Paetorius, in press). The simplicity of the assessment may account for this higher level of true score variance.

**Design-based research.** Due to the fact that this study utilized DBR but was not designed to study DBR, the contributions that this study can make to this field of literature are limited. Nonetheless, there are two contributions worth noting. First, this study provides an example of how DBR can be used in the development of a routine of practice. The DBR approach to research design has shown itself, in this study, to be well suited to the process of creating both the interventions and assessments that are necessary components of the identification of routines of practice. Second, this study provides an example of the potential that exists in DBR studies to advance knowledge in a field. DBR studies are characterized by the dual goals of developing theoretical understanding and developing effective interventions. On one hand, the findings reported in this study are specific to a certain sample of preservice teachers and are not meant to be generalized

to the larger population of preservice teachers. On the other hand, because the study begins with a conceptual framework for the process of EBP and explicitly links the conceptual framework to the intervention, other researchers who wish to address the process of linking research to practice can build upon the explication of this framework.

### **Practice Implications**

The previous section describes the research implications of this study and is meant to address the interests of researchers. In this section, I address the interests of teacher educators and teachers as I describe the practice implications of the study.

This study provides a roadmap for teacher educators who wish to prepare teachers to be able to engage in the process of linking research to practice. The five-step framework provides teacher educators with a conceptual understanding of the complexity that is involved in the process of EBP. The findings of the study, particularly those from the fifth iteration, provide teacher educators with an intervention, the means to assess the intervention, and an example of outcomes from a sample of preservice teachers who experienced the intervention. The resources necessary to conduct the intervention as well as the performance assessment are well documented. Due to the fact that DBR research privileges ecological validity, only typically available resources were utilized in this study (e.g., no external funding was necessary). Furthermore, the documentation of the iterative process of the development of the intervention provides teacher educators with an example of how to continue to refine the intervention and its assessment. In particular, the study provides an example of how general or content-specific teacher educators (versus measurement experts) can leverage the power of generalizability theory

to better understand the conditions that contribute to performance assessment score reliability.

It has been noted that there is a lack of consensus among teacher educators as to the curriculum and pedagogy of teacher preparation programs (Morris & Hiebert, 2009). Often teacher educators have the difficult task of compressing vast volumes of information into a limited number of courses and field experiences. Instead of viewing the routine of practice established in this study as yet one more skill that must be crammed into teacher preparation coursework, I suggest that teacher educators think of this routine of practice as a way to empower future teachers to engage in self-directed professional development throughout their future careers.

Although teacher educators cannot possibly teach preservice teachers all the effective teaching practices that they will need to reach each of their future students, teacher educators can equip preservice teachers with the capacity to find research to address future classroom problems, to appraise that research, to create a plan to implement the research, to modify that plan to meet the needs of specific students, and to evaluate that implementation. In other words, teacher educators can prepare preservice teachers to engage in the process of EBP.

Prior to this study's explication of a process of EBP in the context of teaching, it has been educational researchers, and sometimes administrators, who have analyzed the practices of teachers and classified these practices according to their fidelity to research findings (e.g., Antil et al., 1998, Klinger et al., 2003). I suggest that this study has the potential to contribute to a shifting of the balance of power from the researchers who approve or disapprove of teaching practices back to the teachers who are, themselves,

equipped to assess their own use of practices supported by research. Far from encouraging teachers to engage in scripted, anti-inquiry practices, as some have suggested (Anderson & Herr, 2011), training on the process of EBP equips teachers to apply research findings in sophisticated, tailored ways that are designed to meet the needs of specific students in specific classrooms. Social work researchers have disseminated their work on the process of EBP directly to the social work practitioners (e.g., Rubin, 2008) and I envision that a similar dissemination to teachers may contribute directly to their professional practices.

### **Directions for Future Research**

Just as the findings from this DBR study contribute new knowledge to the literature base, they also generate numerous questions for future research. Here I address four prioritized avenues for further study. First, McKenny and Reeves (2013) emphasized the importance of evaluating DBR studies by looking at the extent to which practice was improved and the problem was solved due to the findings from the study. As it stands, this study demonstrated an improvement in one sample of preservice teachers' abilities to engage in the process of EBP. Documentation of the learning of the EBP process is the first necessary step to understanding the impact upon teaching practice and to addressing the problem of the disconnection between education research and teaching practice. Now future research is needed to explore the extent to which this learning impacts or does not impact the future teaching practices of the learners. Key research questions will be: (1) To what extent do the residents who experienced the learning intervention choose to engage in the EBP process during their inservice teaching? and (2) Does teacher engagement with the process of EBP improve student achievement? This last question



poses significant methodological challenges that have been yet to be overcome in research on EBP in health related fields (e.g., Evidence-Based Medical Working Group, 1992; Hatala & Guyatt, 2002). Nevertheless it is a question of great importance.

Second, now that this DBR study has produced both an intervention and a means of assessment to support that intervention, a multi-site experimental (random assignment to intervention or control group) or quasi-experimental study would increase generalizability and provide evidence that it is the intervention that causes growth in self-reported and observed performance of EBP skills.

Third, the findings from this study indicated that participants developed a lower level of proficiency in modifying and evaluating research implementation as compared to their proficiency in creating an implementation plan. Furthermore, the participants were unaware of this disparity. Therefore, future research might focus on strengthening the support and scaffolding available for these skills. In addition, future research could explore the effect of providing learners with greater field-based feedback on their use of these skills in the classroom.

Finally, additional research is needed to understand why and how some participants experienced more growth as indicated from the reliable change index (RCI) scores for the Familiarity with the Process of Evidence-Based Practice in Education Scale. Future research could explore the extent to which the RCI scores are related or unrelated to future use of the process of EBP. If scores are related to such use, cut off scores that indicate reliable change in EBP behavior could be established for the use of the RCI with the Familiarity with the Process of Evidence-Based Practice in Education Scale.

## **Conclusion**

This DBR study was designed in response to two problems. The first problem, and the one of ultimate consequence, is that there is both lack of mathematics achievement and lack of equity in mathematics achievement in the United States. The second problem is that education research, a potential resource that would enable teachers to facilitate more equitable learning environments, has had little impact upon classroom teaching.

I adapted a conceptual framework from social work, another practitioner-based field, as a model that provided the foundation for an intervention to equip preservice teachers with the skills necessary to engage in the complex process of applying research to practice. Over the course of five iterations, I refined the intervention and generated assessments to support and evaluate the intervention. The study findings documented the self-reported and observed skills of the participants who experienced the intervention as well as the score reliability for the standardized performance assessment.

Nearly a decade ago, the National Council of Teachers of Mathematics (NCTM) expressed a strong commitment to linking research to practice:

Research impacting mathematics education is increasingly important in the decision-making that characterizes the day-to-day work of school district personnel, classroom teachers, and policymakers. In response to these needs, NCTM has adopted as a major goal the linking of research and practice (Gutstein et al., 2005, p.99)

This study represents a commitment to advance this goal forward by equipping future teachers with specific skills that are necessary components to the complex process of

applying research to practice. The findings of this study indicate that merely informing teachers about specific practices supported by research or even modeling these practices grossly underestimates the complexity of the process involved in linking research to practice.

It is my hope that researchers, teacher educators, and teachers themselves will use these findings as a way to empower teachers with the knowledge and skills they need to leverage education research to meet the needs of the diverse students whom they encounter. Far from reducing teachers to the role of technocrats, the process of EBP places equipped teachers at the center of the classroom decision-making process—their rightful place and the place that gives great hope for student learning.

## References

- Adams, K., Matto, H. C., & Lecroy, C. W. (2009). Limitations of evidence-based practice for social work education: Unpacking the complexity. *Journal of Social Work Education, 45*, 165-186.
- AERA/APA/NCME (1999). *Standards for educational and psychological testing*. Washington: APA.
- Anderson, G. L., & Herr, K. (2011). Scaling up “evidence-based” practices for teachers is a profitable but discredited paradigm. *Educational Researcher, 40*, 287-289.
- Anderson, T., & Shattuck, J. (2012). Design-based research: A decade of progress in education research? *Educational Researcher, 41*, 16-25. doi: 10.3102/0013189X11428813
- Anagnostopoulos, D., Smith, E. R., & Basmadjian, K. G. (2007). Bridging the university-school divide: Horizontal expertise and the “two-world” pitfall. *Journal of Teacher Education, 58*, 138-152. doi: 10.1177/0022487106297841
- Allinder, R. M. (2001). Effects of modeling on attitudes and perceptions of preservice special education teachers. *Teacher Education and Special Education, 24*, 362-372. doi: 10.1177/088840640102400410
- Antil, L. R., Jenkins, J. R., & Wayne, S. K. (1998). Cooperative learning: Prevalence, conceptualizations, and the relation between research and practice. *American Educational Research Journal, 35*, 419-454.

- Arbaugh, F., Herbel-Eisenmann, B., Ramirez, N., Knuth, E., Kranendonk, H., & Quander, J. R. (2010). *Linking research and practice: The NCTM research agenda conference report*. Reston, VA: National Council of Teachers of Mathematics.
- Aud, S., Fox, M., & Kewal Ramani, A. (2010). *Status and trends in the education of racial and ethnic groups* (NCES 2010-015). U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office.
- Baartman, L. K. J., Bastiaens, T. J., Kirschner, P. A., & Van der Vleuten, C. P. M. (2007). Evaluation assessment quality in competence-based education: A qualitative comparison of two frameworks. *Educational Research Review*, 2, 114-129.
- Baker, E. L. (2008). *Empirically determining the instructional sensitivity of an accountability test*. Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing.
- Ball, A. F. (2011). American Educational Research Association 2012 annual meeting call for submissions. *Educational Researcher*, 40, 198-220. doi: 10.3102/0013189X11411282
- Ball, D. L., & Cohen, D. K. (1999). Developing practice, developing practitioners: Toward a practice-based theory of professional education. In G. Sykes and L. Darling-Hammond (Eds.), *Teaching as the learning profession: Handbook of policy and practice* (pp. 3-32). San Francisco: Jossey Bass.

- Ball, D. L., Sleep, L., Boerst, T., & Bass, H. (2009). Combining the development of practice and the practice of development in teacher education. *Elementary School Journal, 10*, 458-476.
- Bain, A., Lancaster, J., Zundans, L., & Parkes, R. J. (2009). Embedding evidence-based practice in preservice teacher preparation. *Teacher Education and Special Education, 32*, 215-225. doi: 10.1177/0888406409339999
- Barab, S., & Squire, K. (2004). Design-based research: Putting a stake in the ground. *The Journal of the Learning Sciences, 13*(1), 1-14.
- Bilsker, D., & Goldner, E. M. (2000). Teaching evidence-based practice in mental health. *Research on Social Work Practice, 10*(5), 664-669.
- Boardman, A. G., Arguelles, M. E., Vaughn, S., Hughes, M. T., & Klingner, J. (2005). Special education teachers' views of research-based practices. *Journal of Special Education, 39*, 168-180.
- Boerst, T., Sleep, L., Ball, D. L., & Bass, H. (2011). Preparing teachers to lead mathematics discussions. *Teachers College Record, 113*, 2844-2877.
- Boerst, T., Sleep, L., Shaughnessy, M., & Ball, D. (2012). Assessing preservice teachers' enacted mathematics teacher practice through simulation. Paper presented at the research pre-session of the Annual Meeting of the National Council of Teachers of Mathematics, Philadelphia, PA.
- Boulet, J. R., Murray D., & Kras, J. (2003). Reliability and validity of a simulation based acute care skills assessment for medical students and residents. *Anesthesiology, 91*, 1270-80.

- Brennan, R. L. (2000). Performance assessment from the perspective of generalizability theory. *Applied Psychological Measurement, 24*, 339-353.
- Brown, A. L. (1992). Design experiments: Theoretical and methodological challenges in creating complex interventions in classroom settings. *The Journal of the Learning Sciences, 2*(2), 141-178.
- Cady, J. (2006). Developing mathematics teachers: The transition from preservice to experienced teacher. *The Journal of Educational Research, 99*(5), 295-306. doi: 10.3200/JOER.99.5.295-306
- Cheng, G. (2003). Educational workshop improved information-seeking skills, knowledge, attitudes and the search outcome of hospital clinicians: a randomised controlled trial. *Health Information Library Journal, 20*, 22–33.
- Choudry, N. K., Fletcher, R. H., & Soumerai, S. (2005). Systematic review: The relationship between clinical experience and quality of healthcare. *Annals of Internal Medicine, 142*, 260-273.
- Civil, M. (2011). Mathematics education, language policy, and English language learners. In W. F. Tate, K. D. King, & C. Rousseau Anderson (Eds.), *Disrupting tradition: Research and practice pathways in mathematics education* (pp. 77-92). Reston, VA: National Council of Teachers of Mathematics.
- Cobb, P., Confrey, J., diSessa, A., Lehrer, R., & Schauble, L. (2003). Design experiments in educational research. *Educational Researcher, 32* (1), 9-13.
- Cochran-Smith, M., & Zeichner, K. M. (Eds.). (2005). *Studying teacher education: The report of the AERA Panel on Research and Teacher Education*. Mahwah, NJ: Erlbaum.

- Collins, A. (1992). Toward a design science of education. In E. Scanlon & T. O'Shea (Eds.), *New directions in educational technology* (pp. 15-22). Berlin: Springer-Verlag.
- Collins, A., Joseph, D., & Bielaczyc, K. (2004). Design research: Theoretical and methodological issues. *The Journal of the Learning Sciences, 13* (1), 15-42.
- Confrey, J. (2006). The evolution of design studies as methodology. In K. Sawyer (Ed.), *Cambridge handbook of the learning sciences*. (pp. 135-152). New York: Cambridge University Press.
- Conigliaro, R. L., & Stratton, T. D. (2010). Assessing the quality of clinical teaching: a preliminary study. *Medical Education, 44*, 379–386.
- Cook, B., Landrum, T., Tankersley, M., & Kauffman, J. (2003). Bringing research to bear on practice: Effecting evidence-based instruction for students with emotional or behavioral disorders. *Education and Treatment of Children, 26*(4), 345-361.
- Coomarasamy, A., & Khan K. S. (2004). What is the evidence that post-graduate teaching in evidence-based medicine changes anything? A systematic review. *British Medical Journal, 329*, 1017-1022.
- Cordingley, P. (2008). Research and evidence-informed practice: Focusing on practice and practitioners. *Cambridge Journal of Education, 38*(1), 37-52.
- Crespo, S., Oslund, J. A., & Parks, A. N. (2011). Imagining mathematics teaching practice: Prospective teachers generate representations of a class discussion. *ZDM: International Journal of Mathematics Education, 43*(1), 119-131.



- Cronbach, L. J. (1969). Validation of educational measures. *Proceedings of the 1969 Invitational Conference on Testing Problems: Toward a theory of achievement measurement*. Princeton, New Jersey: Educational Testing Service.
- Cronbach, L. J., Gleser, G. C., Nanda, H., & Rajaratnam, N. (1972). *The dependability of behavioral measurements: Theory of generalizability scores and profiles*. New York: Wiley.
- De Geest, E. (2010). Roles of research in the professional development of mathematics teachers. In M. Joubert & P. Andrews, (Eds.) *Proceedings of the British Congress for Mathematics Education*, 67-72. London: BSRLM.
- Del Carlo, D., Hinkhouse, H., & Isbell, L. (2010). Developing a reflective practitioner through the connection between educational research and reflective practices. *Journal of Science Education and Technology*, 19(1), 58-68. doi: 10.1007/s10956-009-9178-y
- Design-based Research Collective. (2003). Design-based research: An emerging paradigm for educational inquiry. *Educational Researcher*, 32(1), 5-8.
- Dewey, J. (1904/1965). The relation of theory to practice in education. Reprinted in M. Borrowman (Ed.), *Teacher education in America: A documentary history* (pp. 140-171). New York: Teachers College Press.
- diSessa, A. A., & Cobb, P. (2004). Ontological innovation and the role of theory in design experiments. *The Journal of the Learning Sciences*, 13(1), 77-103.

- Dizon, J. M. R., Grimmer-Somers, K. A., & Kumar, S. (2012). Current evidence on evidence-based practice training in allied health: A systematic review of the literature. *International Journal of Evidence-Based Healthcare, 10*, 347-360. doi: 10.1111/j.1744-1609.2012.00295.x
- Edmond, T., Meglvern, D., Williams, C, Rochman, E., & Howard, M. (2006). Integrating evidence-based practice and social work field education. *Journal of Social Work Education, 42*, 377-396.
- English, L. D. (2003). Reconciling theory, research, and practice: A models and modeling perspective. *Educational Studies in Mathematics, 54*, 225-248.
- EU Commission. (2007). The relationship between research, policy and practice in education and training. Unpublished working paper for the European Commission, May.  
[ec.europa.eu/dgs/education\\_culture/publ/pdf/.../sec1098\\_en.pdf](http://ec.europa.eu/dgs/education_culture/publ/pdf/.../sec1098_en.pdf)
- Everett, S A., Luera, G. R., & Otto, C. A. (2008). Pre-service elementary teachers bridge the gap between research and practice. *International Journal of Science and Mathematical Education, 6*(1), 1-17.
- Evidence-Based Medicine Working Group. (1992). Evidence-based medicine: A new approach to teaching the practice of medicine. *Journal of the American Medical Association, 268*, 2420-2425.
- Flores-Mateo, G., & Argimon, J. M. (2007). Evidence based practice in post-graduate healthcare education: A systematic review. *BMC Health Services Research, 7*. doi: 10.1186/1472-6963-7-119.

- Franke, M., Grossman, P., Hatch, T., Richert, A., & Schultz, K. (2006, April). Using representations of practice in teacher education. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.
- Franklin, C. (2007). Teaching evidence-based practices: Strategies for implementation: A response to Mullen et al. and Proctor. *Research on Social Work Practice, 17*, 592-602.
- Florida Department of Education. (2012). Statewide comparison report. Retrieved from <http://fcad.fldoe.org/mediapacket/2012/pdf/MathStatewideComparison.pdf>.
- Fuchs, L. W., & Fuchs, D. (2001). Principles for sustaining research-based practice in the schools: A case study. *Focus on Exceptional Children, 33*(6), 1-14.
- Ghousseini, H. N. (2008). *Learning with routines: Preservice teachers learning to lead classroom mathematics discussions*. Ph.D. dissertation, the University of Michigan.
- Gibbs, L., & Gambrill, E. (2002). Evidence-based practice: Counter-arguments to objections. *Research on Social Work Practice, 12*, 452-476.
- Gira, E., Kessler, M., & Poertner, J. (2004). Influencing social workers to use research evidence in practice: Lessons from medicine and the allied health professions. *Research on Social Work Practice, 14*, 68-79.
- Gitlin, A., Barlow, L., Burbanks, M. D., Kauchak, D. & Stevens, T. (1999). Pre-service teachers' thinking on research: Implications for inquiry oriented teacher education. *Teaching and Teacher Education, 15*, 753-770.

- Gonzales, P., Williams, T., Jocelyn, L., Roey, S., Kastberg, D., and Brenwald, S. (2008). *Highlights from TIMSS 2007: Mathematics and science achievement of U.S. fourth- and eighth-grade students in an international context* (NCES 2009–001 Revised). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.
- Goos, M. (2008). Researcher-teacher relationships in mathematics education. In M. Goos, R. Brown, & K. Makar (Eds.), *Proceedings of the Annual Conference of the Mathematics Education Research Group of Australasia* (pp. 227-233). Sydney: MERGA.
- Gordon, R., Kane, T. J., & Staiger, D. O. (2006). Identifying effective teachers using performance on the job (Discussion paper 2006-01). Washington, DC: Brookings Institution.
- Gore, J. M., & Gitlin, A. D. (2004). [RE]Visioning the academic–teacher divide: power and knowledge in the educational community. *Teachers and Teaching: Theory and Practice*, 10(1), 35-58. doi: 10.1080/13540600320000170918
- Gotham, H. J. (2006). Advancing the implementation of evidence-based practices into clinical practice: How do we get there from here? *Professional Psychology: Research and Practice*, 37, 606-613.
- Greenwood, C R., & Maheady, L. (2001). Are future teachers aware of the gap between research and practice and what should they know? *Teacher Education and Special Education*, 24, 333-347. doi: 10.1177/088840640102400407

- Grossman, P., Compton, C., Igra, D., Ronfeldt, M., Shahan, E., & Williamson, P. (2009). Teaching practice: A cross-professional perspective. *Teachers College Record*, *111*, 2055-2100.
- Grossman P., & McDonald M. (2008). Back to the future: Directions for research in teaching and teacher education. *American Educational Research Journal*, *45*, 184–205.
- Gulikers, J. T., Baartman, L. K., Biemans, H. J. (2010). Facilitating evaluations of innovative, competence-based assessments: Creating understanding and involving multiple stakeholders. *Evaluation and Program Planning*, *33*, 120-127.  
doi:10.1016/j.evalprogplan.2009.07.002
- Gutstein, E., Fey, J. T., Heid, M. K., DeLoach-Johnson, I., Middleton, J. A., Larson, M., Dougherty, B., & Tunis, H. (2005). Equity in school mathematics education: How can research contribute? *Journal for Research in Mathematics Education*, *36*, 92-100.
- Harrison, L J., Dunn. M., & Coome, K. (2006). Making research relevant in preservice early childhood teacher education. *Journal of Early Childhood Teacher Education*, *27*, 217-229. doi: 10.1080/10901020600843434
- Hatala, R., & Guyatt, G. (2002). Evaluating the teaching of evidence-based medicine. *Journal of the American Medical Association*, *288*, 1110-1112.
- Herbst, P. (2011). *Explaining concepts and propositions*. Deep Blue at the University of Michigan. <http://hdl.handle.net/2027.42/84658>.

- Herbst, P., Aaron, W., Beida, K., & Moore-Russo, D. (2012). *Representations of mathematics teaching and their use in transforming teacher education: Contributions to a pedagogical framework*. Unpublished manuscript.
- Herbst, P. & Chieu, V. M. (2011). Depict: A tool to represent classroom scenarios technical report. *Deep Blue at the University of Michigan*.  
<http://hdl.handle.net/2027.42/87949>
- Hiebert, J., & Morris, A. K. (2009). Building a knowledge base for teacher education: An experience in K–8 mathematics teacher preparation. *Elementary School Journal*, 109, 475– 490.
- Hill, H. C., Charalambous, C. Y., & Kraft, M. A. (2012). When rater reliability is not enough: Teacher observation systems and a case for the generalizability study. *Educational Researcher*, 41(2), 56-64.
- Howard, M. O., Allen-Meares, P., & Ruffolo, M. C. (2007). Teaching evidence-based practice: Strategic and pedagogical recommendations for schools of social work. *Research on Social Work Practice*, 17, 561-568. doi: 0.1177/1049731507300191
- Jacobson N. S., Follette, W. C. & Revenstorf, D. (1984). Psychotherapy outcome research: Methods for reporting variability and evaluating clinical significance. *Behavior Therapy*, 15, 336–52.
- Jacobson, N. S. & Traux, P. (1991). Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology*, 59, 12–19.

- Jönsson, A. & Mattsson, M. (2011). Assessing teacher competency during practicum. In M. Mattsson, T.V. Eilertsen & D. Rorrison, (Eds.), *A practicum turn in teacher education* (pp. 169-186). Rotterdam: SensePublishers.
- Jonsson, A., & Svingby, G. (2007). The use of scoring rubrics: Reliability, validity and educational consequences. *Educational Research Review*, 2, 130–44.
- Joshi, R. M., Binks, E., Hougen, M., Dahlgren, M. E., Ocker-Dean, E., & Smith, D. L. (2009). Why elementary teachers might be inadequately prepared to teach reading. *Journal of Learning Disabilities*, 42, 392-402. doi: 10.1177/0022219409338736
- Kazemi, E., Franke, M. & Lampert, M. (2009). Developing pedagogies in teacher education to support novice teachers' ability to enact ambitious instruction. In R. Hunter, B. Bicknell, & T. Burgess (Eds.), *Crossing divides: Proceedings of the 32nd annual conference of the Mathematics Education Research Group of Australasia* (pp. 11-29). Palmerston North, NZ: MERGA.
- Kennedy, M. M. (1997). How teachers connect research and practice. *Midwestern Educational Researcher*, 10(1), 25-29.
- Khan, K. S., & Coomarasamy, A. (2006). A hierarchy of effective teaching and learning to acquire competence in evidenced-based medicine. *BMC Medical Education*, 6, 59, 1-7. doi: 10.1186/1472-6920-6-59.
- Klingner, J. K., Ahwee, S., Pilonieta, P., & Menendez, R. (2003). Barriers and facilitators in scaling up research-based practices. *Exceptional Children*, 69, 411–429.

- Kohler, F., Henning, J. E., & Usma-Wilches, J. (2008). Preparing preservice teachers to make instructional decisions: An examination of data from the teacher work sample. *Teaching and Teacher Education, 24*, 2108-2117.  
doi:10.1016/j.tate.2008.04.002
- Lamb, J., Cooper, T., & Warren, E. (2007). Combining teaching experiments and professional learning: Conflicts between research and teacher outcomes. *Mathematics Education Research Journal, 19*(3), 73-92.
- Lampert, M. (2010). Learning teaching in, from, and for practice: What do we mean? *Journal of Teacher Education, 61*, 21-34.
- Lampert, M., Beasley, H., Ghouseini, H., Kazemi, E., & Franke, M. (2010). Using designed instructional activities to enable novices to manage ambitious mathematics teaching. In M. Stein & L. Kucan (Eds.), *Instructional explanations in the disciplines* (pp. 129–141). New York: Springer.
- Lampert, M., & Graziani, F. (2009). Instructional activities as a tool for teachers' and teacher educators' learning in and for ambitious practice. *Elementary School Journal, 109*, 491-509.
- Landrum, T. J., Cook, B. G., Tankersley, M., & Fitzgerald, S. (2007). Teacher perceptions of the useability of intervention information from personal versus data-based sources. *Education and Treatment of Children, 30*(4), 27-42.
- Linn, R. L., Baker, E. L., & Dunbar, S. B. (1991). Complex performance-based assessment: Expectations and validation criteria. *Educational Researcher, 8*, 15-21.



- Lubienski, S. T., & Gutierrez, R. (2008). Bridging the gaps in perspectives on equity in mathematics education. *The Journal for Research in Mathematics Education*, 7, 365-371.
- Lubinski, C. A., Otto, A. D., & Rich, B. S. (1996, October). *Developing an understanding of K-8 novice teachers: Content knowledge, pedagogical content knowledge, and attitudes*. Paper presented at the annual meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education, Panama City, Florida.
- Lugtenberg, M., Burgers, J. S., & Westert, G. P. (2009). Effects of evidence-based clinical practice guidelines on quality of care: a systematic review. *Quality Safety Healthcare*, 18, 385-392. doi: 10.1136/qshc.2008.028043
- Martin, W. G., Strutchens, M. E., Stuckwisch, S., & Qazi, M. (2011). Transforming East Alabama Mathematics (TEAM-Math): Promoting systemic change in schools and universities. In W. F. Tate, K. D. King, & C. Rousseau Anderson (Eds.), *Disrupting tradition: Research and practice pathways in mathematics education* (pp. 105-118). Reston, VA: National Council of Teachers of Mathematics.
- McDonnough, J T. & Matkins, J. J. (2010). The role of field experience in elementary preservice teachers. *School Science and Mathematics*, 110, 13-23.  
doi: 10.1111/j.1949-8594.2009.00003.x
- McCluskey, A., & Lovarini, M. (2005). Providing education on evidence-based practice improved knowledge but did not change behavior: A before and after study. *BMC Medical Education*, 5, 1-12.

- Messick, S. (1981). Evidence and ethics in the evaluation of tests. *Educational Researcher, 10*, 9-20.
- Messick, S. (1984). The psychology of educational measurement. *Educational Measurement, 21*, 215-237.
- Messick, S. (1989). Validity. In R.L. Linn (Ed.), *Educational measurement* (3rd Ed.) (pp. 13-103). New York: American Council on Education/Macmillan.
- Messick, S. (1994). The interplay of evidence and consequences in the validation of performance assessments. *Educational Researcher, 23*(2), 13-23.
- Miller, D.N., George, M.P., & Fogt, J.B. (2005). Establishing and sustaining research-based practices at Centennial School: A descriptive case study of systemic change. *Psychology in the Schools, 42*, 553–567. doi: 10.1002/pits.20091
- Michie, S., Johnston, M., Abraham, C., Lawton, R., Parker, D., & Walker, A. (2005). Making psychological theory useful for implementing evidence based practice: A consensus approach. *Quality and Safety in Health Care, 14*, 26-33.
- Montecinos, C., Rittershausen, S., Solis, M. C., Contreras, I. & Contreras, C. (2010). Standards-based performance assessment for the evaluation of student teachers: A consequential validity study. *Asia - Pacific Journal of Teacher Education, 38*, 285-300.
- Morris, A., & Hiebert, J. (2009). Introduction: Building knowledge bases and improving systems of practice. *Elementary School Journal, 109*, 429–441
- Moss, P. A. (2011). Analyzing the teaching of professional practice. *Teachers College Record, 113*, 2878-2896.

- Mullen, E. J. (2004). Facilitating practitioner use of evidence-based practice. In A. R. Roberts & K. R. Yeager (Eds.), *Evidence-based practice manual: Research and outcome measures in health and human services* (pp. 205-210). New York: Oxford University Press.
- Mullen, E. J. (2006). Facilitating practitioner use of evidence-based practice. In A. R. Roberts & K. R. Yeager (Eds.), *Foundations of evidence-based social work practice* (pp. 152-159). New York: Oxford University Press.
- Mullen, E. J., Bellamy, J. L., Bledsoe, S. E., & Francois, J. J. (2007). Teaching evidence-based practice. *Research on Social Work Practice, 17*, 574-582.
- Mullen, E. J., Bledsoe, S. E., & Bellamy, J. L. (2008). Implementing evidence-based social work practice. *Research on Social Work Practice, 18*, 325-338.
- Mullen, E. J., Shlonsky, A., Bledsoe, S. E., & Bellamy, J. L. (2005). From concept to implementation: Challenges facing evidence based social work. *Evidence & Policy, 1*(1), 61-84.
- No Child Left Behind Act (Elementary and Secondary Education Act), US, 2002.  
Available at: <http://www.ed.gov/policy/elsec/leg/esea02/index.html>
- Nye, B., Konstantopoulos, S., & Hedges, L. V. (2004). How large are teacher effects? *Educational Evaluation and Policy Analysis, 26*, 237-257.
- OECD. (2007). *Knowledge management, evidence in education: Linking research and policy*. Paris, OECD.
- OECD. (2009). *Creating effective teaching and learning environments: First results from TALIS*, OECD Publishing. doi: 10.1787/9789264068780-en.

- Okhremtchouk, I., Seiki, S., Gilliland, B., Ateh, C., Wallace, M., & Kato, A. (2009).  
Voices in pre-service teachers: Perspectives on the performance assessment for  
California teachers (PACT). *Issues in Teacher Education, 18*, 39-62
- Parrish, D. E., & Rubin, A. (2011). An effective model for continuing education training  
in evidence-based practice. *Research on Social Work Practice, 21*, 77-87. doi:  
10.1177/1049731509359187
- Peck, C., Galluci, C., & Sloan, T., (2010). Negotiating implementation of high-stakes  
performance assessment policies in teacher education: From compliance to  
inquiry. *Journal of Teacher Education, 61*, 451-463.
- Pianta, R. C. & Hamre, B. K. (2009). Conceptualization, measurement, and improvement  
of classroom processes: Standardized observation can leverage capacity.  
*Educational Researcher, 38*, 109-119. DOI: 10.3102/0013189X09332374
- Pignotti, M., & Thyer, B. A. (2012). Novel unsupported and empirically supported  
therapies: Patterns of usage among licensed clinical social workers. *Behavioral  
and Cognitive Psychotherapy, 40*, 331-349. doi: 10.1017/S135246581100052X
- Potari, D., Sakonidis, H., Chatzigoula, R., & Manaridis, A. (2010). Teachers' and  
researchers' collaboration in analyzing mathematics teaching: A context for  
professional reflection and development. *Journal of Mathematics Teacher  
Education, 13*, 473-485. doi:10.1007/s10857-010-9161-9
- Praetorius, A. K., Lenske, G., & Helmke, A. (in press). Observer ratings of instructional  
quality: Do they fulfill what they promise? *Learning and Instruction.*

- Proctor, E. K. (2007). Implementing evidence-based practice in social work education: Principles, strategies, and partnerships. *Research on Social Work Practice, 17*, 583-591.
- Raines, J. (2004). Evidence-based practice in school social work: A process in perspective. *Children and Schools, 26*, 71-85.
- Ratcliffe, M., Bartholomew, H., Hames, V., Hind, A., Leach, J., Millar, R., & Osborne, J. (2005). Evidence-based practice in science education: The researcher-user interface. *Research Papers in Education, 2*, 169-186. doi: 10.1080/0267152050007836
- Riggs, M. L., Verdi, M. P., & Arlin, P. K. (2009). A local evaluation of the reliability, validity, and procedural adequacy of the teacher performance assessment exam for teaching credential candidates. *Issues in Teacher Education, 18*, 13-38.
- Rivkin, S. G., Hanushek, E. A., & Kain, J. F. (2005). Teachers, schools, and academic achievement. *Econometrica, 73*(2), 417–458.
- Rowan, B., Correnti, R., & Miller, R. J. (2002). What large-scale, survey research tells us about teacher effects on student achievement: Insights from the ‘‘Prospects’’ study of elementary schools. *Teachers College Record, 104*(8), 1525–1567.
- Rubin, A. (2007). Improving the teaching of evidence-based practice: Introduction to the Special Issue. *Research on Social Work Practice, 15*, 541-547. DOI: 10.1177/1049731507300145
- Rubin, A. (2008). *Practitioner’s guide to using research for evidence based practice*. Hoboken, NJ: John Wiley & Sons.

- Rubin, A., & Parrish, D. (2007). Views of evidence-based practice among faculty in master of social work programs: A national survey. *Research on Social Work Practice, 17*, 110-122.
- Rubin, A., & Parrish, D. (2011). An effective model for continuing education training in evidence-based practice. *Research on Social Work Practice, 21*, 77-87. doi: 10.1177/1049731509359187
- Sackett, D. L., & Rosenberg, W.M. (1995). The need for evidence-based medicine. *Journal of the Royal Society of Medicine, 88*, 620-624.
- Sackett, D. L., Rosenberg, W. M. C., Gray, J. A. M., Haynes, R. B., & Richardson, W. S. (1996). Evidence based medicine: What it is and what it isn't. *British Medical Journal, 312*, 71-72.
- Sackett, D. L., Straus, S. E., Richardson, W. S., Rosenberg, W. M. C., & Haynes, R. B. (2000). *Evidence-based medicine: How to practice and teach EBM* (2nd ed.). New York: Churchill Livingstone.
- Sanders, W. L., & Rivers, J. C. (1996). *Cumulative and residual effects of teachers on future student academic achievement*. Knoxville: University of Tennessee, Value-Added Research and Assessment Center.
- Sandholtz, J. H., & Shea, L. M. (2012). Predicting performance: A comparison of university supervisors' predictions and teacher candidates' scores on a teaching performance assessment. *Journal of Teacher Education, 63*(1), 39-50.
- Satterfield, J. M., Spring, B., Brownson, R. C., Mullen, E. J., Newhouse, R. P., & Whitlock, E. P. (2009). Toward a transdisciplinary model of evidence-based practice. *Milbank Quarterly, 87*(2), 368-390.

- Schmidt, W. H. (2011). Building bridges between research and the worlds of policy and practice: Lessons learned from PROM/SE and TIMSS. In W. F. Tate, K. D. King, & C. Rousseau Anderson (Eds.), *Disrupting tradition: Research and practice pathways in mathematics education* (pp. 45-52). Reston, VA: National Council of Teachers of Mathematics.
- Schuster, M. A., McGlynn, E. A., & Brook, R. H. (1998). How good is the quality of health care in the United States? *Milbank Quarterly*, 76, 517-563.
- Shaneyfelt, T. et al. (2006). Instruments for evaluating education in evidence-based practice: a systematic review. *The Journal of the American Medical Association*, 296, 1116–1127.
- Shavelson, R. J., Baxter, G. P., & Pine, J. (1991). Performance assessments in science. *Applied Measurement in Education*, 4(4), 347-362.
- Shavelson, R. J., & Towne, L. (Eds.). (2002). *Scientific research in education*. Washington, DC: National Academy Press.
- Shavelson, R. J., & Webb, N. M. (1991). *Generalizability theory: A primer*. Thousand Oaks, CA: Sage.
- Shlonsky, A., & Gibbs, L. (2004). Will the real evidence-based practice please stand up? Teaching the process of evidence-based practice to the helping professions. *Brief Treatment and Crisis Intervention*, 4, 137-153.

- Siegler, R., Carpenter, T., Fennell, F., Geary, D., Lewis, J., Okamoto, Y., Thompson, L., & Wray, J. (2010). *Developing effective fractions instruction for kindergarten through 8th grade: A practice guide* (NCEE #2010-4039). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. Retrieved from [whatworks.ed.gov/publications/practiceguides](http://whatworks.ed.gov/publications/practiceguides).
- Sinnema, C., Sewell, A., & Milligan, A. (2011). Evidence-informed collaborative inquiry for improving teaching and learning. *Asia - Pacific Journal of Teacher Education*, 39, 247-261. doi: 10.1080/1359866X.2011.597050
- Snyder, J. (2009). Taking stock of performance assessments in teaching. *Issues in Teacher Education*, 18(1), 7-11.
- Stigler, J. W., Hiebert, J. (2009). Closing the teaching gap. *Phi Delta Kappan*, 91(3), 32-37.
- Thyer, B. A. (2006). What is evidence-based practice? In A. R. Roberts & K. R. Yeager (Eds.), *Foundations of social work practice* (pp. 35-46). New York: Oxford University Press.
- Umbeck, L. (2011). Navigating classroom change. *Mathematics Teaching in the Middle School*, 17, 89-95.
- U.S. Department of Education (2010). *ESEA Blueprint for Reform*. Washington, D.C.: Office of Planning, Evaluation and Policy Development.



- Vanderlinde, R., & van Braak, J. (2010). The gap between educational research and practice: Views of teachers, school leaders, intermediaries and researchers. *British Educational Research Journal*, 36, 299-316.  
doi.org/10.1080/01411920902919257
- Wentworth, N., Erickson, L. B., Lawrence, B., Popham, J. A., & Korth, B. (2009). A paradigm shift toward evidence-based clinical practice: Developing a performance assessment. *Studies in Educational Evaluation*, 35(1), 16–20.
- Wise E. A. (2004). Methods for analyzing psychotherapy outcomes: a review of clinical significance, reliable change, and recommendations for future directions. *Journal of Personality Assessment*, 82, 50–59.
- Woodward, J., Beckmann, S., Driscoll, M., Franke, M., Herzig, P., Jitendra, A., Koedinger, K. R., & Ogbuehi, P. (2012). *Improving mathematical problem solving in grades 4 through 8: A practice guide* (NCEE 2012-4055). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. Retrieved from [http://ies.ed.gov/ncee/wwc/publications\\_reviews.aspx#pubsearch/](http://ies.ed.gov/ncee/wwc/publications_reviews.aspx#pubsearch/).
- Wright, S. P., Horn, S. P., & Sanders, W. L. (1997). Teacher and classroom context effects on student achievement: Implications for teacher evaluation. *Journal of Personnel Evaluation in Education*, 1(1), 57–67.
- Zeuli, J. S. (1994). How do teachers understand research when they read it? *Teaching and Teacher Education*, 10, 39-55.

## Appendices

## **Appendix A**

### Education Research Project (Iteration 1)

By this point in the semester, we have learned a great deal about *what* to teach and *how* to teach in K-6 math. At the same time, there is still so much more to learn! It would be impossible to learn everything in one or two semesters of math methods; as educators we commit to being lifelong learners. The purpose of this project is to help to equip you to continue to learn about the best ways to teach elementary school math.

There are 5 skills that help teachers to apply research to practice.

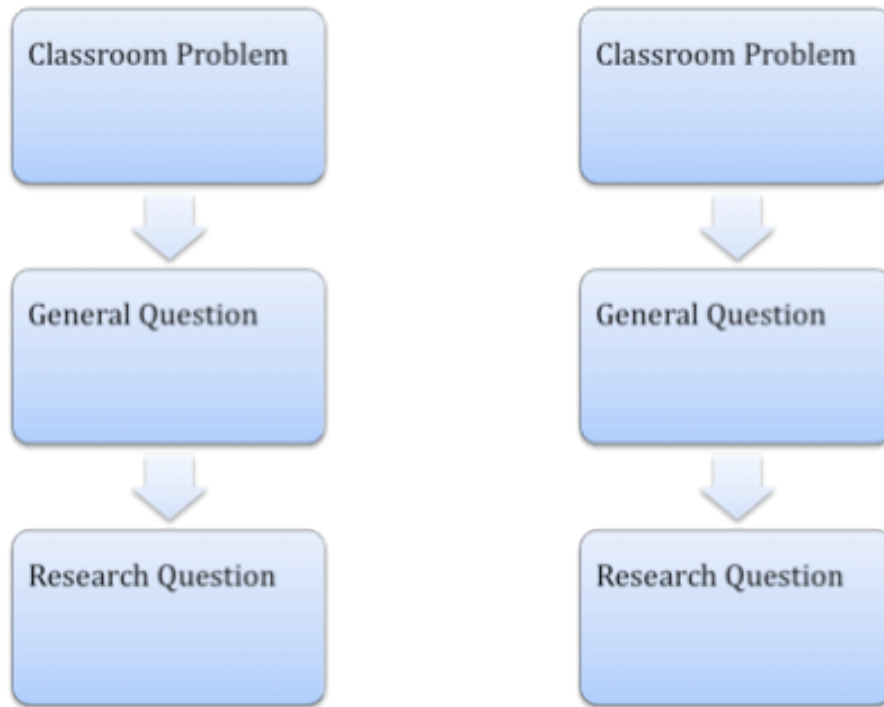
- I. Ask a research question
- II. Find research articles and read them
- III. Synthesize and analyze the articles
- IV. Apply research to the classroom
- V. Participate on a research team

You can work on this project on your own or with a partner. Partners will turn in one project together. You will complete one sheet on each of these 5 skills.

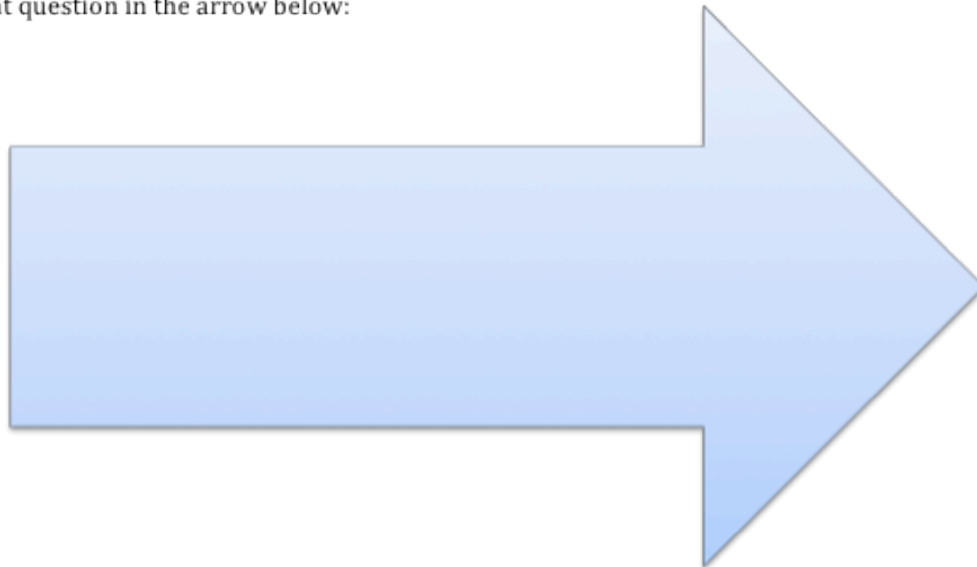
Together the 5 sheets make up a research portfolio. Each sheet is worth 20 points. Please print out all 5 sheets so that you can work on them as time allows in class.

You will also need to attach your research article to this packet.

### Ask a Research Question



Choose ONE of the research questions above that you would like to explore in more detail. Write out that question in the arrow below:



Score: /20

Find and Read Research Articles

After thinking about my question, these are the keywords that I will use in my article search:

Seach Terms:

There are the names of 3 databases that I will search:


This is the citation for one relevant **empirical (data-based) article** that I found. Please use APA 6<sup>th</sup> edition format for this citation.

Author, A. (2010). I am very happy to tell you my story. *The Best Journal Ever Written*,4, 314-332.

Citation:

Attach your paper to this packet now.

Score: /20

Synthesize and Analyze

Normally one would synthesize and analyze the results of several studies. However, for this brief assignment, you will simply summarize your one article. Please summarize in no less than 5 and no more than 10 sentences:

Please circle the best answer:

Was this article easy to read?

1                      2                      3                      4  
Very hard to read.    Somewhat hard.    Somewhat easy.    Very easy.

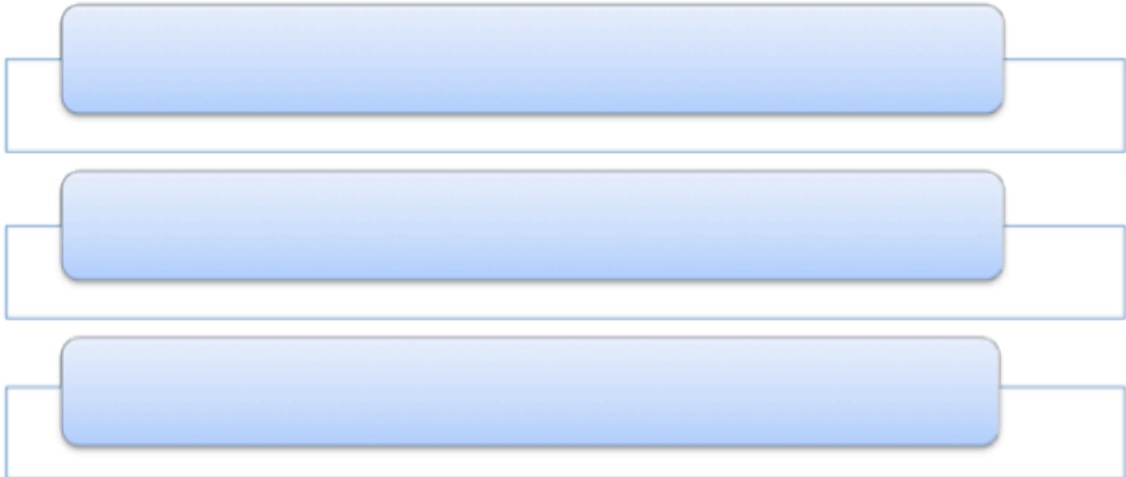
Would it be relatively easy and straightforward to apply this idea in the classroom?

1                      2                      3                      4  
Very hard to apply.    Somewhat hard.    Somewhat easy.    Very easy.

Score: /20

Apply Research to the Classroom:

What are 3 questions you would consider when thinking about applying information gained from a research study into your own classroom:



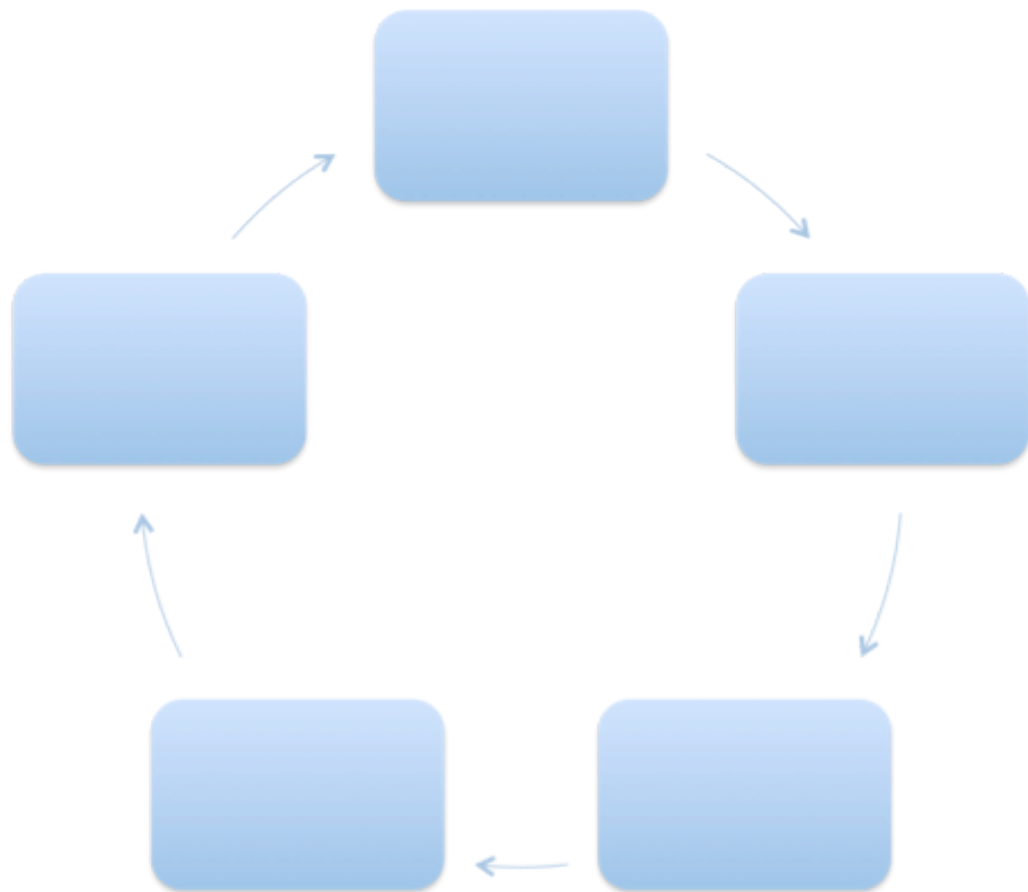
Briefly describe (4-8 sentences) how you personally would apply the results of your research study to a particular grade level.

Score: /20

## Research Participation

### I. The Research Cycle.

Please put the following steps in the appropriate location on the research cycle.



1. Collect Data
2. Define a Problem
3. Make Conclusions
4. Ask a Research Question (Hypothesis)
5. Analyze Data

### II. Research Interests

Let's say that a future principal asks you to be on a research team that is conducting research in the classroom. Given that you must do this for your job, what are 3 areas of K-6 education that you are interested in and about which you might like to conduct research.

- 1.
- 2.
- 3.

Score: /20



## Appendix B

### Information Literacy Questionnaire Pilot Version (Iteration 1)

**Name:**

**Date:**

1. How many times here at USF has a librarian visited one of your classes, OR have you been to the library to receive group instruction on the use of the library and its resources?
  - a. Never
  - b. 1 time
  - c. 2 times
  - d. 3 or more times
2. How many times have you communicated one-on-one with a librarian to get assistance in an information search (face-to-face, email, or chat)?
  - a. Never
  - b. 1 time
  - c. 2 times
  - d. 3 or more times
3. How many semesters have you COMPLETED within the college of Education at USF?
  - a. 0
  - b. 1
  - c. 2
  - d. 3 or more
4. At what point are you with education internships?
  - a. Have not yet taken an internship
  - b. Have begun Level 1 internship this semester
  - c. Completed Level 1 internship prior to this semester
  - d. Have begun Level 2 internship this semester
  - e. Have completed Level 2 internship prior to this semester
5. Have you ever been required to find scholarly articles as references for a required paper/project in an EDUCATION course here at USF?
  - a. Yes
  - b. No
  - c. I don't know
6. How much experience do you think that you have with READING education research?
  - a. No experience
  - b. A little experience
  - c. A great deal of experience
7. As far as you know, are teachers required by law to use teaching practices that are supported by education research?
  - a. Yes
  - b. No
8. When you think about your future teaching career, how likely do you think you are to use the following resources to help inform your teaching? Please rank on a scale of 1 to 5.

	Not at all likely		Somewhat likely		Very Likely
A. Advice from fellow teachers	1	2	3	4	5
B. Advice from administration	1	2	3	4	5
C. Google searches	1	2	3	4	5

D. Scholarly research journal articles	1	2	3	4	5
E. Practitioner journal articles	1	2	3	4	5
F. Textbooks from USF education courses	1	2	3	4	5

9. When on the USF library home page, what would you consider the best place to begin researching a topic in teaching and/or learning?
  - a. Browse through journals using the e-journals link
  - b. Browse through items under the “Digital Collections” link
  - c. Click on “guides” and “subject” to find the education subject guide
  - d. Click on “Databases by title” link
10. You’d like to browse through current issues of *Educational Review*. What is the best, most efficient strategy to locate current articles in this journal?
  - a. Go to the library and browse the second floor in the periodicals room.
  - b. Search the online library catalog to determine if the library subscribes to the journal either online or in print.
  - c. Google the name of the journal and see if the full text of the articles are available for free.
  - d. Search in ERIC for the names of well-known education researchers to see if any have published recently in *Educational Review*. If so, then follow the links to the full articles.
11. When you are in an article database and you do not have the full text available to you in that database, you can get to the full text by clicking on:
  - a. The RefWorks link
  - b. The "Save, Print, E-Mail" link
  - c. The "Findit@USF" link
  - d. The "Search Tools" link
12. What is the difference between subject heading (or descriptor) and keyword searching when using a database?
  - a. Keywords are more difficult to use.
  - b. Subject headings/descriptors are more convenient to use.
  - c. Keywords provide less noise in the results.
  - d. Subject headings offer more relevant results.
13. You have to write a paper about why some children won’t compete when playing games. Which type of specialized database—in addition to an education database— might have relevant scholarly articles?
  - a. History
  - b. Psychology
  - c. Philosophy
  - d. Literature

14. Suppose you want to extract from a periodical database all articles which contain references both to “homework” as well as to “mathematics” (in the same article). Select the expression below that uses the Boolean operator you should use to link the two parts of the search statement?
- Homework ALSO Mathematics
  - Homework OR Mathematics
  - Homework WHILE Mathematics
  - Homework AND Mathematics
15. A keyword search in the library catalog retrieves more than 800 hits. What would be the next best step to focus your search?
- 800 hits is not too many. Look through all of them.
  - Look at the first ten hits and choose the most relevant materials.
  - Try the search over again with fewer terms.
  - Try the search over again with added terms.
16. What is “peer review”?
- A system of review carried out by a committee of congressional delegates.
  - A process, prior to publication, for checking that academic articles have been examined by other researchers in the field.
  - A process for guaranteeing that articles are 100 percent true prior to publication.
  - A process for reviewing research material using multiple, microscopic lenses.
17. What is the purpose of an abstract?
- It gives an abstract understanding of a complex topic.
  - It provides details on the methods used to obtain the research data.
  - It provides important background information on the author who wrote the paper.
  - It provides a brief summary of the most important points in a paper.
18. If a periodical article has a methods and a results section, it is most likely a(n)
- Empirical article
  - Literature review article
  - Theoretical article
  - Position paper
19. Which of the following is characteristic of practitioner articles?
- Articles include detailed descriptions of the research methods and results.
  - Articles are theoretical and only describe abstract concepts.
  - Articles are written for university professors.
  - Articles are focused on application of ideas and trends in the profession.
20. The results of education research create a preponderance of evidence when:
- The results from one study provide evidence that causes the reader to ponder the meaning.
  - The results from several studies contradict each other; thus causing educators to ponder whether they should use the research in their classrooms or not.
  - The results from several, well-designed studies agree with each other, and any contradictory evidence is less supported.

- d. The results from one study show that there is no statistically significant effect of the education method being studied.
21. You know that there is a gap in education research when
- There are no newspaper or popular magazine articles on the topic of interest.
  - You cannot find the information you need from a Google search
  - No scholarly articles have been written on the topic of interest.
  - There is a gap in the citation for a scholarly journal.
22. Which of the following questions could be answered by a research study carried out in an elementary school?
- Who is better at math, 3<sup>rd</sup> grade boys or 3<sup>rd</sup> grade girls?
  - Why aren't sixth grade students as interested in math as they used to be?
  - Is 10 minutes of math homework for 2nd grade students in school XYZ associated with increased FCAT scores as compared with 2nd students with no homework in school XYZ?
  - Are elementary school teachers more interested in teaching reading than teaching math?
23. Please match the following websites with the information that is best obtained on each:
- |                                     |  |
|-------------------------------------|--|
| a. _____ERIC                        |  |
| b. _____NCES                        | 1. scholarly articles                                    |
| c. _____What Works<br>Clearinghouse | 2. evidence-based practices<br>3. educational statistics |
24. Please order these steps from 1(beginning) to 5 (ending) to reflect the most common, logical progression for education research.
- |                        |                      |
|------------------------|----------------------|
| _____ Collect Data     | _____ Ask a Research |
| _____ Define a Problem | Question             |
| _____ Make Conclusions | _____ Analyze Data   |
25. Let's say that you want to try out a teaching strategy that has been shown to be highly effective in a research study. The research article tells you about the strategy, but you have to decide how to apply that strategy in your own classroom. Circle any of the following factors that you would take into consideration when thinking about how to apply the research to your practice:
- The age of your students
  - The ethnicities & cultures of your students
  - The classroom time that you can allocate to the new strategy
  - The achievement level of your students
  - Opportunities for professional development in this strategy
  - Potential conflicts between your district/school policies and new strategy
  - Classroom supplies needed to support this new strategy

## Appendix C

### Original Rubric for Evaluation of First Four Tasks from the Education Research Project (Iteration 1)

1. Was the research question related to the identified problem?	0 = No 1 = Weakly Related 2 = Related 3 = Strongly Related
2. Was the research question manageable?	B = Broad M = Manageable N = Narrow
3. Did student find article that addressed research Q?	0 = No 1 = Weakly relevant 2 = Relevant 3 = Strongly Relevant
4. Did search terms cover all major concepts in question?	0 = No 1 = Yes
5. How was the student's use of search terms?	0 = Inadequate 1 = Barely adequate 2 = Adequate 3 = Skilled

## Appendix D

### Codebook for Grounded Theory Study (Iteration 1)

**Codebook- Codes are listed in alphabetic order.**

**Beliefs override research**

Expresses a belief in such a way that it would take priority over research findings

---

**Believes that she already uses research findings=**

Makes a statement that indicates that she already has been implementing the research

---

**Clearly defined implementation plan**

Expresses a defined plan for implementing research

---

**Concerning statement**

Participant makes a comment, perhaps even peripheral to main point that is of serious concern. For example, the participant may indicate that working with ELL students is a burden or may make a pedagogical mistake.

---

**Conducts research**

Plans to conduct research. Might be as a means to test veracity of original research.  
Might be as a way to evaluate implementation

---

**Connects implementation to stance on teaching and learning**

Is able to provide a WHY or PURPOSE (a particular stance on teaching or learning) to using the research in her practice.

---

**Considers local conditions**

Expresses consideration for local conditions. This consideration may influence the course of implementation.

---

**Does not define expected outcome**

Does not express what she hopes to see happen as a result of implementing the research

---

**Does not identify essential research findings**

Does not express an understanding of WHAT the research was saying. Does not express key elements of an intervention.

---

**Does not identify implementation barrier**

Fails to mention a significant barrier to the implementation that has been suggested.

---

**Global goal but no implementation plan**

Expresses an overall goal for the implementation of the research, but does not provide a plan for how the research would be implemented

---

**Grandiose implementation plan**

Implementation plan is not realistic. Perhaps participant tries to fix deep, pervasive problems with an action plan that superficially addresses problem.

---

**Identifies essential research findings**

Expresses understanding of central message or components of the research

---

**Identifies implementation barrier**

Identifies (and perhaps ameliorates) a potential barrier to the implementation plan that has been described.

---

**Identifies potential support to implementation**

Identifies something that can be leveraged to benefit the implementation of the research idea.

---

**Ignorance of research process**

Makes a statement about conducting research that shows ignorance regarding research process and what are logical expectations of results from research process

---

**Illogical connection between research and implementation**

There is no logical relationship between the research that was presented or suggested and the plan for implementation.

---

**Illogical prediction of result of implementation**

Participant predicts an outcome

---

**Implementation plan strongly related to research**

Provides clear link between research and plan for implementation

---

**Implementation plan weakly related to research**

There is only a weak connection between research as stated and implementation plan.

---

**Lacks connection to stance on teaching and learning**

Does not connect plan for implementation to deeper values that are guided by her stance on teaching and learning. May not express understanding for the reason why she is implementing the research.

---

**Lacks connection to teaching & learning goals**

Suggests action to be taken, but that action is not tied to a stance on teaching and learning. May suggest carrying out an action that has no clear purpose for enhancing the learning process.

---

**Lacks understanding of a concept central to the research**

It becomes clear that the participant has a blind spot in an area of related research that hinders her understanding of the research or its application.

---

---

**No implementation plan**

Provides no plan for implementing the research

---

**No plan to evaluate implementation**

Participants give no concrete indication of how they will evaluate the implementation.

---

**Non-problematized view of research**

Thinks of research in simple and absolutistic terms. For example, "This piece of research 'proves' such and such."

---

**Over-simplification of complex problem**

Suggests a simple course of action that is unlikely to produce the intended results because the simple action does not address the complexity of a problem.

---

**Parent role**

Makes a comment about the role of parent(s)

---

**Passive stance**

The participant abdicates the power and responsibility that she has to make change in the classroom.

---

**Plan to evaluate implementation**

Participants give specific details as to how they will evaluate the implementation of their research.

---

**Plan to share results of implementation**

Participant expresses plan to disseminate findings from implementation.

---

**Poorly defined implementation plan**

Does express a plan for implementation but either it is vague or consists of isolated steps that are presented without a unifying purpose.

---

**Recognizes need for evaluation but has no plan**

Participant expresses the need for evaluation of implementation but does not suggest how this might be accomplished.

---

**Recognizes need for iteration**

Expresses awareness that implementation is an evolving process. This includes making plans to alter implementation based on feedback from initial actions.

---

**Student role**

Makes a comment about the student's role

---

**Takes critical stance toward research**

Has a questioning stance toward research. Can be a useful stance when thinking

---



about how to apply research.

---

**Unclear connection between research and implementation**

It is not clear how the implementation that the participant suggests comes from the research that she cited.

---

**Unrealistic expectation for teacher**

Provides an expectation of a teacher that appears to be unreasonable (perhaps naive about an aspect of teachers' work).

---

**Unrealistic expectation for implementation results**

Participant expects an outcome from implementation that is not likely.

---

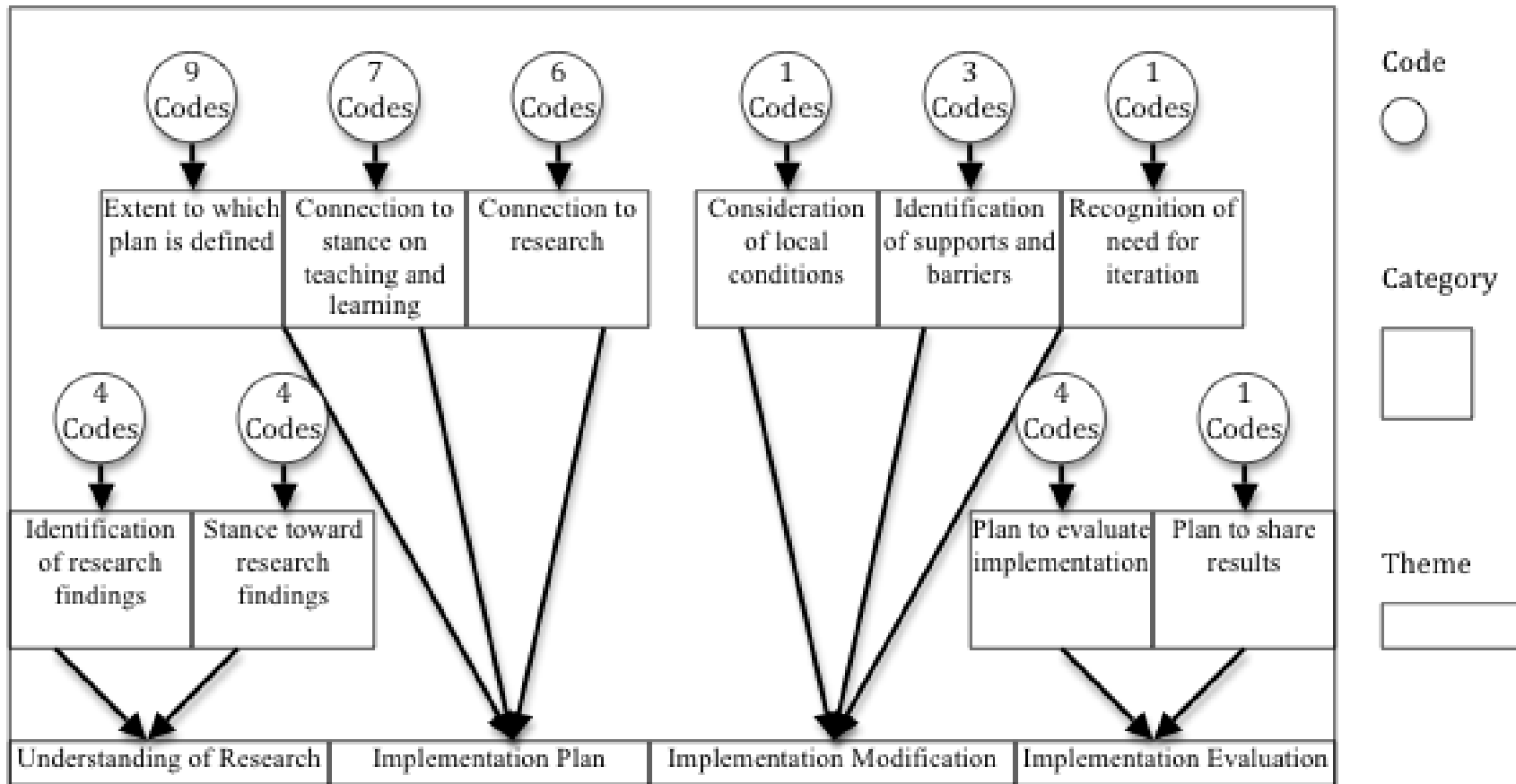
**Wants to control parent/student behaviors**

Makes a comment that directly or indirectly implies that they can control what parent/students think or do. Perhaps she thinks that she can "change" parents.

---

### Appendix E

Relationships among Codes, Categories and Themes from Grounded Theory Analysis of Research Application Essays



## Appendix F

### Information Literacy Questionnaire Short Form (Iteration 2-5)

Name: \_\_\_\_\_

Date: \_\_\_\_\_

1. How many times here at USF has a librarian visited one of your classes, OR have you been to the library to receive group instruction on the use of the library and its resources?
  - a. Never
  - b. 1 time
  - c. 2 times
  - d. 3 or more times
2. How many times have you communicated one-on-one with a librarian to get assistance in an information search (face-to-face, email, or chat)?
  - b. Never
  - b. 1 time
  - c. 2 times
  - d. 3 or more times
3. Have you ever been required to find scholarly articles as references for a required paper/project in an EDUCATION course here at USF?
  - c. Yes
  - b. No
  - c. I don't know
4. How much experience do you think that you have with READING education research?
  - d. No experience
  - b. A little experience
  - c. A great deal of experience
5. As far as you know, are teachers required by law to use teaching practices that are supported by education research?
  - a. Yes
  - b. No

6. When you think about your future teaching career, how likely do you think you are to use the following resources to help inform your teaching? Please rank on a scale of 1 to 5.

	Not at all likely	1	2	Somewhat likely	3	4	Very Likely	5
Advice from fellow teachers	1	2	3	4	5			
Advice from administration	1	2	3	4	5			
Google searches	1	2	3	4	5			
Scholarly research journal articles	1	2	3	4	5			
Practitioner journal articles	1	2	3	4	5			
Textbooks from USF education courses	1	2	3	4	5			

7. What is the purpose of an abstract?
  - e. It gives an abstract understanding of a complex topic.
  - f. It provides details on the methods used to obtain the research data.
  - g. It provides important background information on the author who wrote the paper.
  - h. It provides a brief summary of the most important points in a paper.
8. If a periodical article has a methods and a results section, it is most likely a(n)
  - a. Empirical article
  - b. Literature review article
  - c. Theoretical article
  - d. Position paper
9. Which of the following is characteristic of practitioner articles?
  - e. Articles include detailed descriptions of the research methods and results.
  - f. Articles are theoretical and only describe abstract concepts.
  - g. Articles are written for university professors.
  - h. Articles are focused on application of ideas and trends in the profession.
10. Which of the following questions could be answered by a research study carried out in an elementary school?
  - i. Who is better at math, 3<sup>rd</sup> grade boys or 3<sup>rd</sup> grade girls?
  - j. Why aren't sixth grade students as interested in math as they used to be?

- k. Is 10 minutes of math homework for 2nd grade students in school XYZ associated with increased FCAT scores as compared with 2nd students with no homework in school XYZ?
- l. Are elementary school teachers more interested in teaching reading than teaching math

11. Let's say that you want to try out a teaching strategy that has been shown to be highly effective in a research study. The research article tells you about the strategy, but you have to decide how to apply that strategy in your own classroom. Circle any of the following factors that you would take into consideration when thinking about how to apply the research to your practice:

- The age of your students
- The ethnicities & cultures of your students
- Classroom supplies needed to support strategy
- Potential conflicts between your district/school policies and new strategy
- Opportunities for professional development in this strategy
- The classroom time that you can allocate to the new strategy
- The achievement level of your student

## Appendix G

### Interview Protocol for Semi-Structured Interview of Students Who Completed the Education Research Project

#### Semi-structured Interview Protocol

Prior to the interview, the participant will be given the opportunity to read the consent form and to decide if he/she would like to sign and participate or not participate.

1. Thank you for your willingness to be interviewed.
2. I would like to remind you that this conversation is voluntary. You are free to respond or not respond to any questions that I ask.
3. This conversation is for research purposes only and has no impact on academic standing at USF. Your name or identifying information will not be made known. All of your responses will be anonymous.
4. The purpose of this interview is for me to gain an understanding of your experiences and thinking related to those experiences. I would like to engage in a two-way conversation, so you are free to ask questions yourself at any time during this interview.

#### Questions:

- 1 Could you tell me what internship experiences you've had since you completed your education research project? Could you give me 5 adjectives to describe your internship experiences?
- 2 Can you describe the Education Research Project that you completed in MAE4310 or MAE4326? Prompt as needed with questions such as: what was the topic, what was your experience reading the research article, what do you remember about the research article.
- 3 Was there any time during your internship that you thought about the Education Research Project that you completed for MAE4310 or 4326? Please describe the circumstances and what you were thinking. (Follow up to see if they used any information from this project to guide their teaching decisions). Do you feel that your internship experiences have impeded or encouraged the use of your education research?
- 4 Have you discussed education research with your cooperating teacher or another educator in the school(s) in which you've been interning?
  - a. If yes, please describe.
  - b. If no, do you think you might engage in conversations about education research with other teachers? Explain. (Follow up with a question about what besides education research guides teachers' decisions.)
- 5 In your opinion, how important or unimportant is mathematics education research for the classroom teacher? Explain.
- 6 What do you think about this statement: Mathematics education research is too difficult for elementary teachers to understand.
- 7 Is there any other information that you would like to share with me about the education research project or research in general?

## Appendix H

### Rubric for Formative Assessment of the Revised Education Research Project (Iteration 4)

1. Question Formulation	Was the research question related to the identified problem?	0 No response	1 Weakly Related	2 Related	3 Strongly Related
	Was the research question manageable?	0 Too broad or too narrow		1 Manageable	
2. Evidence Search	Did search terms cover all major and minor concepts in the research question	0 No response	1 Missing term for a major concept	2 Terms for major but not minor	3 Terms for major and minor
	Did student choose appropriate databases?	0 Listed 1 or more inappropriate databases		1 Listed 3 appropriate databases	
	Did student choose appropriate articles?	0 No appropriate articles	1 1 appropriate	2 2 appropriate	3 3 appropriate
	Were articles cited in basic APA format (minor format variations ok)	0 None appropriate	1 1 appropriate	2 2 appropriate	3 3 appropriate
3. Critically Appraise Research	Identified essential research findings for the 3 articles	0 No Essential findings	1 Essential findings for 1 article	2 Essential findings for 2 articles	3 Essential findings for 3 articles
	Synthesis creates a coherent discussion of the similarities and differences among the articles.	0 No synthesis	1 Doesn't articulate both similarities and differences	2 Articulates some similarities and differences	3 Fully articulates both similarities and differences
	Synthesis conclusion articulates essential points regarding the topic given the 3 articles	0 No conclusion	1 Does not synthesize results from all 3 articles	2 Articulates some ideas from each article	3 Fully articulates what is known on topic

4. Implement the Intervention	Written plan clearly connects intervention to stance on teaching and learning.	0 Not connected	1 Weakly connected	2 Connected	3 Explicitly & clearly connected
	The implementation plan is clearly defined (actionable steps).	0 Not defined	1 Weakly defined	2 Defined	3 Clearly defined
	The plan is clearly connected to the research.	0 No connection	1 Connection unclear	2 Connected	3 Clearly connected
	The plan gives consideration to local conditions. This includes identification of possible barriers and supports to implementation.	0 No consideration for modifications	1 Minimal consideration for modifications	2 Important modifications considered	3 Extensive thought given to modification
	Plan clearly articulates how the teacher will evaluate the implementation.	0 No plan for evaluation	1 Plan unclear or unsubstantial	2 Adequate plan	3 Clearly articulated plan

## Appendix I

### Research Summaries for Summative Assessment of the Process of Evidence-Based Practice

## Recommendation 2



### **Assist students in monitoring and reflecting on the problem-solving process.**

*Students learn mathematics and solve problems better when they monitor their thinking and problem-solving steps as they solve problems.<sup>45</sup> Monitoring and reflecting during problem solving helps students think about what they are doing and why they are doing it, evaluate the steps they are taking to solve the problem, and connect new concepts to what they already know. The more students reflect on their problem-solving processes, the better their mathematical reasoning—and their ability to apply this reasoning to new situations—will be.<sup>46</sup>*

*In this recommendation, the panel suggests that teachers help students learn to monitor and reflect on their thought process when they solve math problems. While the ultimate goal is for students to monitor and reflect on their own while solving a problem, teachers may need to support students when a new activity or concept is introduced. For instance, a teacher may provide prompts and use them to model monitoring and reflecting as the teacher solves a problem aloud. In addition, a teacher can use what students say as a basis for helping the students improve their monitoring and reflecting. Teachers can use students' ideas to help students understand the problem-solving process.*

#### **Summary of evidence: Strong Evidence**

Several studies with diverse student samples directly tested this recommendation and consistently found positive effects. As a result, the panel determined there was strong evidence to support this recommendation.<sup>47</sup>



## Recommendation 2 *(continued)*

The relevant studies examined students' mathematics achievement in different content areas, including numbers and operations, data analysis and probability, algebra, and geometry. Two studies found that providing students with a task list that identified specific steps to solving problems resulted in better student achievement.<sup>48</sup> Two additional studies found that a self-questioning checklist improved achievement,<sup>49</sup> and in one study, this effect persisted for at least four

months after instruction ended;<sup>50</sup> however, both studies included additional instructional components (visual aids and multiple-strategy instruction) that may have produced the positive results. Similarly, five studies found that student performance improved when teachers modeled a self-questioning process and then asked students to practice it.<sup>51</sup>

The panel identified three suggestions for how to carry out this recommendation.

### How to carry out the recommendation

1. Provide students with a list of prompts to help them monitor and reflect during the problem-solving process.

The prompts that teachers provide can either be questions that students should ask and answer as they solve problems (see Example 6) or task lists that help students complete steps in the problem-solving process (see Example 7).<sup>52</sup> The questions teachers provide should require students to think through the problem-solving process, similar to the way in which task lists guide students through the process. Select a reasonable number of prompts, rather than an exhaustive list, as too many prompts may slow down the problem-solving process or be ignored. Ensure that the prompts help students evaluate their work at each stage of the problem-solving process, from initially reading and understanding the problem, to determining a way to solve the problem, and then to evaluating the appropriateness of the solution given the facts in the problem.<sup>53</sup>

Encourage students to explain and justify their response to each prompt, either orally<sup>54</sup> or in writing.<sup>55</sup> Students can use the prompts when working independently, in small groups,<sup>56</sup> or even when solving problems at a computer.<sup>57</sup> When working in small groups, students can

take turns asking and answering questions or reading each action aloud and responding to it. As they share in small groups, students serve as models for others in their group, allowing all the students to learn from one another. Teachers may wish to post prompts on the board, include them on worksheets,<sup>58</sup> or list them on index cards for students.<sup>59</sup>

When students first use the prompts, they may need help. Teachers can participate in the questioning or refer to tasks in the task list when students work in small groups or during whole-group discussions. If, for example, a student solves a problem incorrectly, ask him what questions he should have asked himself to help him reason out loud, rather than providing him with the correct answer.<sup>60</sup> Alternatively, provide the correct answer, but ask the student to explain why it is right and why his original answer is not.<sup>61</sup> As students become more comfortable with their reasoning abilities and take greater responsibility for monitoring and reflecting during problem solving, teachers can gradually withdraw the amount of support they provide.<sup>62</sup>

## Recommendation 3



### **Help students understand why procedures for computations with fractions make sense.**

*Students are most proficient at applying computational procedures when they understand why those procedures make sense. Although conceptual understanding is foundational for the correct use of procedures, students often are taught computational procedures with fractions without an adequate explanation of how or why the procedures work.*

*Teachers should take the time to provide such explanations and to emphasize how fraction computation procedures transform the fractions in meaningful ways. In other words, they should focus on both conceptual understanding and procedural fluency and should emphasize the connections between them. The panel recommends several practices for developing understanding of computational procedures, including use of visual representations and estimation to reinforce conceptual understanding. Addressing students' misconceptions and setting problems in real-world contexts also can contribute to improved understanding.*

#### **Summary of evidence: Moderate Evidence**

The panel based this recommendation in large part on three well-designed studies that demonstrated the effectiveness of teaching conceptual understanding when developing students' computational skill with fractions.<sup>73</sup> These studies focused on decimals and were relatively small in scale; however, the panel believes that their results, together with extensive evidence showing that meaningful

information is remembered much better than meaningless information, provide persuasive evidence for this recommendation.<sup>74</sup> Additional support for the recommendation comes from four studies that showed a positive relation between conceptual and computational knowledge of fractions.<sup>75</sup>

The studies that contributed to the evidence base for this recommendation used computer-based interventions to examine the link

### Recommendation 3 *continued*

between conceptual knowledge and computational skill with decimals. Sixth-grade students completed three lessons on decimal place value (i.e., conceptual knowledge) and three lessons on addition and subtraction of decimals (i.e., procedural knowledge).<sup>76</sup> Iterating between the two types of lessons improved students' procedural knowledge, compared with teaching all of the conceptual lessons before any of the procedural ones. In another study, 5th- and 6th-grade students practiced locating decimals on a number line using a computer-based game. Dividing the number line into tenths and encouraging students to notice the tenths digit improved 5th- and 6th-grade students' ability to locate decimals on a number line (compared to not providing the prompts).<sup>77</sup>

Research also shows a positive relationship between students' conceptual and procedural knowledge of fractions. That is, children who have above-average conceptual knowledge also tend to have above-average knowledge of computational procedures. Studies of 4th- and 5th-graders and of 7th- and 8th-graders indicated that conceptual knowledge was positively related to computational proficiency after controlling for prior math achievement, arithmetic fluency, working memory, and reading ability.<sup>78</sup> In addition, conceptual knowledge of decimals predicted students' ability to locate decimals on a number line.<sup>79</sup> While these studies show a correlation between conceptual and procedural knowledge, they did not examine the effectiveness of interventions that develop conceptual knowledge to improve procedural knowledge.

The panel also identified evidence that specifically addressed two of the four steps for implementing this recommendation.

**Use of representations.** Evidence identified by the panel supports the recommended

practice of using visual representations and manipulatives during instruction on fraction computation (Step 1). Two well-designed studies found that the use of manipulatives and pictorial representations had a positive effect on computational skill with fractions.<sup>80</sup> One of these studies focused on fraction circles (sets of circles, in which the first is a whole circle, the second is divided in half, the third is divided in thirds, etc.).<sup>81</sup> The other study had students use a variety of manipulatives for learning computational procedures with fractions, including fraction squares and fraction strips.<sup>82</sup> A third study examined the Rational Number Project curriculum, which emphasizes the use of manipulatives as one of many components.<sup>83</sup> The authors of the study reported that the curriculum had a positive effect on fraction computation abilities. However, manipulatives were only one component of this multifaceted curriculum, and the study provided insufficient information for the WWC to complete a review, so the conclusions that can be drawn from the study regarding the role of manipulatives are limited.

**Real-world contexts.** The panel identified evidence related to the use of real-world contexts for improving skill at executing computational procedures with fractions (Step 4).<sup>84</sup> In one of the studies, personalizing problems for 5th- and 6th-grade students improved their ability to solve division problems with fractions.<sup>85</sup> The other study found that posing problems in everyday contexts improved 11- and 12-year-old students' ability to order and compare decimals.<sup>86</sup> Additional studies argued for the use of real-world contexts for teaching procedures for computing with fractions but did not provide rigorous evidence that such instruction causes improvement in fraction computation.<sup>87</sup>

## Appendix J

### Rubric for Scoring of the Summative Standardized Assessment of the Process of Evidence-Based Practice

	Below Expectation	Meets Expectation	Exceeds Expectation
The implementation plan is clearly defined (actionable steps).	1 The plan has broad goals but does not provide specific actions that could be taken by a classroom teacher.	2 The plan has specific steps that a classroom teacher can accomplish, but steps may be difficult to accomplish.	3 The plan has specific steps AND the steps are reasonable for a motivated teacher to accomplish.
The plan is clearly connected to the research.	1 The written plan does not articulate a clear connection to the research.	2 The plan articulates a general connection to the research but the plan does not explicitly state how specific implementation steps reflect research findings.	3 The plan is recognizably connected to the research AND the teacher explicitly connects various parts of the plan to the research findings.
The plan articulates the need to make modifications to research implementation to meet the needs of the particular students in the classroom.	1 Plan makes no mention of the need to modify research implementation according to the particular needs of students in a unique classroom.	2 Plan includes one comment that indicates either a modification for students' needs OR the intention to modify during the implementation process based upon the needs of students in a unique classroom.	3 Plan includes more than one comment that indicate a modification for students' needs AND/OR the intention to modify during the implementation process based upon the needs of students in a unique classroom.
The plan anticipates a barrier to the implementation of the research findings.	1 The plan does not articulate a potential barrier to research implementation.	2 Plan lists one potential barrier to the implementation.	3 Plan lists more than one potential barrier to implementation.
The plan provides a potential remedy/support to anticipated barriers to implementation.	1 The plan does not mention any remedies/supports that can be leveraged to address potential barriers to implementation.	2 The plan articulates one potential remedy/support that can be leveraged to address potential barriers to implementation.	3 The plan articulates more than one potential remedy/support that can be leveraged to address potential barriers to implementation.
Plan clearly articulates how the teacher will evaluate the implementation.	1 There is no plan for how teacher will evaluate effectiveness of the implementation.	2 There is a plan for how to evaluate implementation, but it is brief and without detail.	3 There is a detailed plan for how the teacher will evaluate the implementation.
The plan for evaluation of the implementation could reasonably be carried out by a classroom teacher.	1 There is no plan for how teacher will evaluate effectiveness of the implementation.	2 There is a plan to evaluate implementation but it is either vague (lacks detail) OR would be difficult to accomplish with normal classroom resources.	3 This plan for evaluation is specific and could be used by most classroom teachers.

## Appendix K

### Initial and Final Versions of the Familiarity with the Process of Evidence-Based Practice in Education Scale with Summary Form from Expert Review Committee

#### Initial Familiarity with the Process of Evidence-Based Practice in Education Scale

If you were to enter the classroom as an independent teacher today, how strongly do you agree or disagree that you could perform the following tasks? 1= Strongly disagree, 2= Somewhat disagree, 3= Neutral 4=Somewhat agree, 5= Strongly Agree, and X=I do not know what this means.

	Strongly Disagree	Somewhat Disagree	Neutral	Somewhat Agree	Strongly Agree	I do not know what this means
When I have a problem in classroom teaching, I can formulate a question that can be answered by education research.	1	2	3	4	5	X
Once I have a research question, I know how to look for research articles that address the question.	1	2	3	4	5	X
I can tell if an education article is theoretical or data-based.	1	2	3	4	5	X
I can tell if a data-based article uses quantitative, qualitative or mixed methodology.	1	2	3	4	5	X
I can judge whether a research article provides strong or weak evidence.	1	2	3	4	5	X
I can synthesize the evidence from several research articles.	1	2	3	4	5	X
I know how to create a plan to implement education research findings in my own classroom.	1	2	3	4	5	X
I am able to adapt the research findings to meet the needs of my particular students.	1	2	3	4	5	X
I know how to identify potential barriers that might hinder my attempt to implement education research in my classroom.	1	2	3	4	5	X
I know how to evaluate whether my use of research-based findings has been successful.	1	2	3	4	5	X

### Summary Form from Three-Person Expert Review Panel

	Aligned with the framework for process of EBP as applied to teaching?		Clear?		Single Idea?		Suggestions Concerns Notes
	Yes	No	Yes	No	Yes	No	
							R3: I usually use neither agree nor disagree
When I have a problem in classroom teaching, I can formulate a question that can be answered by education research.	R2 R3	R1	R1 R2 R3		R1 R2 R3		R1: I would suggest rewording as shown to the left. I would imagine EBP for teaching could be broader and extend beyond problems to finding the most effective ways to teach different student populations? I would word this more broadly... R3: Should this be educational research? Is it necessary to specify "education?"
Once I have a research question, I know how to look for research articles that address the question.	R2 R3	R1	R2 R3	R1	R2 R3	R1	R1: A research question would be very different than an educational question? A research question is answered by a research study while an educational or teaching question is posed to rely on the existing research literature to answer that question? Not sure what the right wording would be, but thought education-related question might work? R2: Are "how" and "where" different things here? Probably "how" encompasses "where," so this is probably okay, but might measure more than one skill.
I can tell if an education article is theoretical or data-based.	R1 R3	R2	R1 R2 R3		R1 R2 R3		R2: See above.
I can tell if a data-based article uses quantitative, qualitative or mixed methodology.	R1 R3	R2	R1 R2 R3		R1 R2 R3		

I can judge whether a research article provides strong or weak evidence.	R1 R2 R3		R1 R2 R3		R1 R2 R3		R2: See above.
I can synthesize the evidence from several research articles.	R1 R2 R3		R2 R3	R1	R1 R2 R3		R1:I would clarify a bit more why this is done
I know how to create a plan to implement education research findings in my own classroom.	R2 R3	R1	R1 R2 R3		R1 R2 R3		R1:I would use “best available research”, as this is consistent with EBP.
I am able to adapt the research findings to meet the needs of my particular students.	R2 R3	R1	R1 R2 R3		R1 R2 R3		R1:I would use “best available research”, as this is consistent with EBP.
I know how to identify potential barriers that might hinder my attempt to implement education research in my classroom.	R2 R3	R1	R1 R3	R2?	R1 R3	R2?	R1:I would use “best available research”, as this is consistent with EBP. R2: This is somewhat vague, and could mean many things. Is there a more specific thing you are trying to measure here? I imagine giving an example would be too specific, but in reading this, I didn't really know what the prompt might
I know how to evaluate whether my use of research-based findings has been successful.	R1 R2 R3		R1 R2 R3		R1 R2 R3		R1:Just some suggestions in rewording to make this clearer.

Note: R1 = Reviewer 1, R2 = Reviewer 2, R3 = Reviewer 3

**Final Familiarity with the Process of Evidence-Based Practice in Education Scale**

In the context of classroom teaching, how strongly do you agree or disagree that you could perform the following tasks? 1= Strongly disagree, 2= Somewhat disagree, 3= Neutral 4=Somewhat agree, 5= Strongly Agree, and X=I do not know what this means.

	<b>Strongly Disagree</b>	<b>Somewhat Disagree</b>	<b>Neutral</b>	<b>Somewhat Agree</b>	<b>Strongly Agree</b>	<b>I do not know what this means</b>
I can formulate a question to inform my teaching in the classroom that can be answered by education research.	1	2	3	4	5	X
Once I have an education related question, I know how to find research articles to answer this question.	1	2	3	4	5	X
I can tell if an education article is theoretical or data-based.	1	2	3	4	5	X
I can tell if a data-based article uses quantitative, qualitative or mixed methodology.	1	2	3	4	5	X
I can judge whether a research article provides strong or weak evidence.	1	2	3	4	5	X
I can synthesize the evidence from several research articles to answer my educational or teaching question.	1	2	3	4	5	X
I know how to create a plan to implement the best available education research in my own classroom.	1	2	3	4	5	X
I am able to adapt the best available education research to meet the needs of my particular students.	1	2	3	4	5	X
I know how to identify potential barriers that might hinder my attempt to implement the best available research in my classroom.	1	2	3	4	5	X
I know how to evaluate whether my research-based education plan has been successful within my classroom.	1	2	3	4	5	X



## Appendix L

### Quiz on Basic Information Literacy Skills (Iteration 5)

**Name** Library Resources Quiz

**Description** This quiz is based on some of the material in the content area entitled "Library Resources."

**Instructions** Please review the videos and links provided in the "Getting Started" and "Research Help" folders before you start this quiz.

- **Multiple Choice: In the short video, How to Connect from...**

**Question** In the short video, How to Connect from Off-Campus, what login method is recommended as the BEST?

- Answer**
- A. Signing in through Blackboard and then navigating to the USF Library home page.
  - B. Going through an email link provided by your USF email account.
  - C. Clicking on the "login for full access" link on the top left hand side of the library home page.
  - D. Clicking on the library catalog and then on the button that says "renewals/library account."

**Correct Feedback** Yes, direct login to the library home page is normally the best option. The Blackboard option can sometimes cause difficulty when trying to access our e-books.

**Incorrect Feedback** Sorry but this answer is incorrect.

- **Multiple Choice: With respect to the Ask-a-Librarian...**

**Question** With respect to the *Ask-a-Librarian* services, which method of assistance is **NOT** currently available?

- Answer**
- Chat service
  - Email service
  - Face to face service
  - Skype service

**Correct Feedback** That's right we do not offer Skype service.

**Incorrect Feedback** No, this is a service we do provide.

- **Multiple Choice: According to the video about Ask a...**

**Question** According to the video about *Ask a Librarian* services, when chat services are

not available after regular hours, how can you get help anyway?

**Answer**

- Through email or scheduling an appointment with a librarian.
- By coming into the library and hoping for the best.
- By telephoning the library.
- By using the texting service.

**Correct Feedback** The email service and appointments in advance with staff are your best option if it's after hours when the library is under staffed or when chat is unavailable.

**Incorrect Feedback** Other than email and scheduled appointments, after hours service is very limited through chat, phone or in the library.

- **Multiple Choice: Which type of Ask-a-Librarian...**

**Question** Which type of Ask-a-Librarian assistance is likely to provide the most comprehensive assistance with a research project?

**Answer**

- Chat services
- Consultation services
- Phone services
- Email services

**Correct Feedback** Correct! Consulting with a librarian who is an expert in your major area of study can offer you the most in depth assistance.

**Incorrect Feedback** Incorrect. See the Ask a Librarian Services at USF video in the "Getting Started" folder under "Library Resources."

- **Multiple Choice: Which resource would be a good starting...**

**Question** Which resource would be a good starting point for your research if you wanted to identify a list of education databases or reference resources?

**Answer**

- The library catalog.
- Databases by title list.
- The education subject guide.
- The e-journals link

**Correct Feedback** Yes the education subject guide gives you a list of good starting points for both databases and reference tools related to education, according to the video about course guides and subject guides.

**Incorrect Feedback** The catalog would not be a place that helps you find a list of resources in once place. Databases by title is only helpful if you know what databases you're looking for. Otherwise, you might not find what you need to. Finally the e-journals link is not an efficient way to find either databases or reference titles, nor is it an easy way to look through the literature in education since you'd have to search e-journals title by title.

- **Multiple Choice: When you use a database that is only...**

**Question** When you use a database that is only partially full text, which link would you use to find the full text from a citation that does not have a pdf for the article?

**Answer**

- RefWorks "export" link
- Findit@USF link
- Advanced search link
- Save/Print/Email link

**Correct Feedback** Yes, the Findit@USF button is your friend in terms of navigating you to the full text online.

**Incorrect Feedback** The Findit@USF links will navigate you to the full text. See the video about finding full text articles for review.

- **Multiple Choice: Based on the video about using books...**

**Question** Based on the video about using books vs. articles for a class project, what resources are most useful if you need a broad overview of a subject or want some background information?

**Answer**

- Articles from academic journals.
- Articles from trade journals
- Books
- Blog posts

**Correct Feedback** Books or reference books are the best choice if you are looking for background or overviews about broad topics.

**Incorrect Feedback** While journals seem like a good choice, it's not always a good idea to dive into them if you need an overview or background information. Blog posts are not considered reputable sources, nor are they good places to get overviews. Review the video about books vs. articles for more information.

- **Multiple Choice: According to the video tutorial about...**

**Question** According to the video tutorial about scholarly vs. popular periodicals from Vanderbilt University (see the folder called "Research Help--Beyond the Basics," scholarly journals are considered scholarly because they contain what type of material?

**Answer**

- articles written by reporters for the public
- articles written by experts in the field
- material written for people in a particular trade or profession about new trends
- glossy pictures and job ads

**Correct Feedback** Yes, authorship and authority are an important characteristic of scholarly journals.

**Incorrect Feedback** Sorry, you got this one wrong. You might want to review the material in the tutorial to make sure you get it right when selecting the right sources for class projects in college.

- **Multiple Choice: According to the tutorial from...**

**Question** According to the tutorial from Vanderbilt University, "peer review" in relation to journals or articles can be defined as:

**Answer**

- A process prior to publication where manuscripts are evaluated by other experts in the field
- A system of review carried out by a committee of congressional delegates
- A process for guaranteeing that articles are 100 percent true prior to publication
- A process where consumers rate the value of the journals

**Correct Feedback** Yes. Peer review is all about determining if a manuscript is good enough to publish in a well-ranked publication or not. Normally an editorial board of experts decides that.

**Incorrect Feedback** Incorrect answer. Revisit the tutorial about peer review in the Blackboard folder entitled "Research help--Beyond the basics."

- **Multiple Choice: As mentioned in the tutorial about...**

**Question** As mentioned in the tutorial about scholarly vs. popular periodicals, what is the purpose of an abstract in journal articles?

**Answer**

- It gives an abstract understanding of a complex topic.
- It evaluates the methods used to obtain the research data.
- It provides biographical information about the author who wrote the paper.
- It offers a brief summary of the most important points in the paper.

**Correct Feedback** Correct! Reading abstracts about articles can be a great time-saver when evaluating your sources and looking for the right information.

**Incorrect Feedback** Incorrect answer. Review the Scholarly vs. Peer Review tutorial under "Research Help--Beyond the Basics"

**Question** What is the difference between descriptors (or subjects) and keywords when searching a very large database?

**Answer**

Keywords are more difficult to use

Subject terms/descriptors are more convenient to use

Subject terms/descriptors offer more relevant results

Keywords provide less "noise" in the results

**Correct Feedback** Well done!

**Incorrect  
Feedback**

Incorrect answer. Review the database thesauri tutorial.

## Appendix M

### Rubric for Scoring Education Research Project (Iteration 5)

Rubric					
1. Question	Was the research question related to the identified problem?	0 No response	1 Weakly Related	2 Related	3 Strongly Related
	Was the research question manageable?	0 Too broad or too narrow		1 Manageable	
2. Evidence Search	Did search terms cover all major and minor concepts in the research question	0 No response	1 Missing term for a major concept	2 Terms for major but not minor	3 Terms for major and minor
	Did student choose appropriate databases?	0 Listed 1 or more inappropriate databases		1 Listed 3 appropriate databases	
	Did student choose appropriate articles?	0 No appropriate articles	1 1 appropriate	2 2 appropriate	3 3 appropriate
	Were articles cited in basic APA format (minor format variations ok)	0 None appropriate	1 1 appropriate	2 2 appropriate	3 3 appropriate
3. Critically Appraise Research	Identified essential research findings for the 3 articles	0 No Essential findings	1 Essential findings for 1 article	2 Essential findings for 2 articles	3 Essential findings for 3 articles
	Synthesis creates a coherent discussion of the similarities and differences among the articles.	0 No synthesis	1 Doesn't articulate both similarities and differences	2 Articulates some similarities and differences	3 Fully articulates both similarities and differences
	Synthesis conclusion articulates essential points regarding the topic given the 3 articles	0 No conclusion	1 Does not synthesize results from all 3 articles	2 Articulates some ideas from each article	3 Fully articulates what is known on topic

4. Implement the Intervention (Plan)	Written plan clearly connects intervention to stance on teaching and learning.	0 Not connected	1 Weakly connected	2 Connected	3 Explicitly and clearly connected
	The implementation plan is clearly defined (actionable steps).	0 Not defined	1 Weakly defined	2 Defined	3 Clearly defined
	Plan is clearly connected to the research.	0 No connection	1 Connection unclear	2 Connected	3 Clearly connected
	Plan gives consideration to local conditions: includes identification of possible barriers and supports to implementation.	0 No consideration for modifications	1 Minimal consideration for modifications	2 Important modifications considered	3 Extensive thought given to modification
	Plan clearly articulates how the teacher will evaluate the implementation.	0 No plan for evaluation	1 Plan unclear or unsubstantial	2 Adequate plan	3 Clearly articulated plan
4. Implement the Intervention (Reflection)	The reflection explains how the teaching was connected to the research. (Refl. 1)	0 No connection	1 Connection unclear	2 Connected	3 Clearly connected
	Reflection explains how the teacher modified the research implementation to meet the needs of students (Reflection1)	0 No explanation of modifications	1 Minimal explanation of modifications	2 Important modifications explained	3 Extensive explanation of modifications
5. Evaluate effect on teaching and learning	Articulates effect of implementation on teacher and students. (Reflection 2)	0 No effects given	1 Missing teacher or student	2 Effects of both teacher & student	3 Clearly articulated effects
	Provides evidence of success/lack of success of implementation based on video or artifacts. (Reflection 3)	0 No evidence	1 Weak evidence	2 Adequate evidence	3 Substantial evidence
	Makes reasonable recommendations for future implementation. (Reflection 3)	0 None given	1 No basis for recommendations	2 Basic recommendations	3 Comprehensive recommendations

### Appendix N

Sample Table for Reporting Results from Generalizability Theory Analysis

Source	n	df	Estimated Variance Component	% Variance
Persons(p)	10			
Rater (r)	3			
Task (t)	2			
Occasion (o)	2			
p x r				
p x t				
p x o				
r x t				
r x o				
t x o				
p x r x t				
p x r x o				
p x t x o				
r x t x o				
p x r x t x o, e				



## Appendix O

### Sample Data Setup for Generalizability Theory Analysis

Person	Rater	Task	Occasion	Score
1	1	1	1	#
1	1	1	2	#
1	1	2	1	#
1	1	2	2	#
1	2	1	1	#
1	2	1	2	#
1	2	2	1	#
1	2	2	2	#
1	3	1	1	#
1	3	1	2	#
1	3	2	1	#
1	3	2	2	#

## **Appendix P**

### Standardized Performance Assessment Follow-up Interview Protocol

Please read the article  
Please read your response.

#### **Plan:**

Ask the student to describe the details of the plan. Can the student provide actionable steps if she did not in the written reflection. Ask the student about connection to research. Prompt student to identify connections if she did not identify on assessment.

#### **Modifications:**

Ask student if she sees a potential need for modification.

#### **Barrier:**

Ask student if she sees a potential barrier to implementation.  
Ask student if she sees a remedy/support to counteract the barrier.

#### **Evaluation:**

Ask the student how she would know if the plan worked? See if she acknowledges need for evaluation plan.  
Does the student have a plan to evaluate the implementation? Prompt the student to reflect upon whether or not that plan can reasonably be carried out.

# Appendix Q

## Copyright Permissions

2/13/13

University of South Florida Mail - (no subject)



Sarah VanIngen <vaningen@mail.usf.edu>

---

**(no subject)**

1 message

---

**Myrna Jacobs** <mjacobs@nctm.org>  
To: vaningen@mail.usf.edu

Fri, Jan 11, 2013 at 9:45 AM

Dear Ms. VanIngen,

You are welcome to reproduce the figure you reference below. Best of luck on your dissertation.

Sincerely,

**Myrna Jacobs** | *Publications Manager*

National Council of Teachers of Mathematics  
1906 Association Drive | Reston, VA 20191

703.620.9840 x 2192 | fax 703.715.9536

[www.nctm.org](http://www.nctm.org)

**[Browse NCTM's newest titles!](#)**

**From:** Sarah VanIngen [<mailto:vaningen@mail.usf.edu>]

**Sent:** Thursday, January 10, 2013 12:38 PM

**To:** NCTM

**Subject:** NCTM Web Site Feedback

Referring page: <http://www.nctm.org/publications/content.aspx?id=522>

I am requesting permission to reproduce (and cite appropriately) Figure 3 from Pg 33 from Linking Research to Practice: The NCTM Research Conference Report. I would like to reproduce this figure for my dissertation.

This page on your website asks me to contact the Copyright Clearance Center. However, this center does not have the record of this document. Could you provide me with the contact of the person/institution that is able to give permission for the use of material from the Linking Research to Practice Report?

Thank you very much!

<https://mail.google.com/mail/u/0/?ui=2&ik=5ab1783b96&view=pt&q=NCTM&size=20&pnr=100&pdr=50&search=apps&th=13c2a12ffe1b5c87>

1/2



Sarah VanIngen &lt;vaningen@mail.usf.edu&gt;

---

**Request to reproduce table in dissertation**

2 messages

---

**Sarah VanIngen** <vaningen@mail.usf.edu>  
To: permissionsUS@wiley.com

Fri, Feb 8, 2013 at 12:29 PM

To Whom It May Concern:

I am requesting permission to reproduce in my dissertation Table 5 (Our Revised EBP Model) from the following source:

**Title:** Toward a Transdisciplinary Model  
of Evidence-Based Practice

**Author:** JASON M. SATTERFIELD, BONNIE  
SPRING, ROSS C.  
BROWNSON, EDWARD J.  
MULLEN, ROBIN P.  
NEWHOUSE, BARBARA B.  
WALKER, EVELYN P. WHITLOCK

**Publication:** Milbank Quarterly

**Publisher:** John Wiley and Sons

**Date:** Jun 4, 2009

Thank you very much for your assistance!  
Sarah van Ingen

—

Ms. Sarah van Ingen

Instructor and Doctoral Candidate  
Mathematics Education  
Department of Secondary Education  
USF College of Education

---

**Permission Requests - UK** <permissionsuk@wiley.com>  
To: "vaningen@mail.usf.edu" <vaningen@mail.usf.edu>

Thu, Feb 21, 2013 at 6:43 AM

Dear Ms. Sarah van Ingen,

Thank you for your email request.

<https://mail.google.com/mail/u/0/?ui=2&ik=5ab1783b96&view=pt&search=inbox&th=13cbada716c0ef8b>

1/3

Permission is granted for you to use the material requested for your thesis/dissertation subject to the usual acknowledgements and on the understanding that you will reapply for permission if you wish to distribute or publish your thesis/dissertation commercially.

Permission is granted solely for use in conjunction with the thesis, and the article may not be posted online separately.

Any third party material is expressly excluded from this permission. If any material appears within the article with credit to another source, authorisation from that source must be obtained.

Best Wishes,

Verity Butler

Permissions Co-ordinator

Wiley

The Atrium, Southern Gate  
Chichester, PO19 8SQ

UK

[www.wiley.com](http://www.wiley.com)

[vbutler@wiley.com](mailto:vbutler@wiley.com)

**WILEY**

John Wiley & Sons Ltd is a private limited company registered in England with registered number 641132.  
Registered office address: The Atrium, Southern Gate, Chichester, West Sussex, United Kingdom. PO19 8SQ.

---

**From:** Goldweber, Paulette - Hoboken **On Behalf Of** Permissions - US

<https://mail.google.com/mail/u/0/?ui=2&ik=5ab1783b96&view=pt&search=inbox&th=13cbada716c0ef8b>

2/3