

To the Graduate Council:

I am submitting a dissertation written by Rebecca E. Pennington entitled “Measuring the Effects of an Instructional Scaffolding Intervention on Reflective Thinking in Elementary Preservice Teacher Developmental Portfolios.” I have examined the final electronic copy of this dissertation and recommend that it be accepted in partial fulfillment of the requirements for the Doctor of Education, with a major in Learning and Leadership.

---

Dr. M. D. Roblyer, Chairperson

We have read this dissertation  
and recommend its acceptance.

---

Dr. Lloyd Davis

---

Dr. Deborah McAllister

---

Dr. Bryan Stewart

Accepted for the Graduate Council:

---

Dr. Stephanie Bellar  
Interim Dean of the Graduate School

MEASURING THE EFFECTS OF  
AN INSTRUCTIONAL SCAFFOLDING INTERVENTION ON  
REFLECTIVE THINKING IN ELEMENTARY PRESERVICE TEACHER  
DEVELOPMENTAL PORTFOLIOS

A Dissertation  
Presented for the  
Doctor of Education Degree  
The University of Tennessee at Chattanooga

Rebecca E. Pennington  
July 2010

..... **Copyright © Rebecca E. Pennington, 2010**

All rights reserved

## **Dedication**

This dissertation is dedicated to my loving husband, Aubrey Tucker Pennington, Jr., whose unflagging support and faith in me made this work possible. I am thankful for his faithful love, care, sense of humor, and intellectual challenge.

## **Acknowledgments**

It is my pleasure to acknowledge the many people who have supported and sustained me through the work of earning the degree of Doctor of Education in Learning and Leadership from The University of Tennessee at Chattanooga. I would like to thank my dissertation committee for their invaluable assistance throughout this process. Dr. Deborah McAllister's attention to detail and Dr. Bryan Stewart's practical wisdom contributed greatly to this project. I am especially grateful to Dr. Lloyd Davis for sharing his deep knowledge of methodology and data analysis.

I am indebted to my chair, Dr. Margaret (M. D.) Roblyer, for her boundless energy and for shaping my thoughts during my coursework and during the dissertation process. Throughout each step in this journey Dr. Roblyer challenged me, cheered for me, and championed my work with expertise and humor. I am grateful for her kind leadership that made writing the dissertation a joyous adventure.

I would like to thank my colleagues at Covenant College for their continued encouragement: Dr. Phillip Horton for his statistical expertise and service as sounding board, Dr. Daphne Haddad and Jane Stewart for their calm assurances, and Mrs. Amy Bagby and Dr. Jack Beckman for their hours of tireless work assisting me with data collection.

Finally, I wish to express my deep gratitude to my family, whose unending support sustained me and made completion of this degree possible. My husband, Tucker Pennington, never wavered in his love and belief that I would finish. My daughters, Hannah and Aubrianna, patiently endured the challenges of living with a mom who was often more distracted doctoral student than parent.

## Abstract

This study was designed to determine whether teacher portfolios can be validly and reliably assessed, to investigate the effect of an instructional tool on increasing the level of reflective thinking in elementary preservice teachers' portfolios, and to find whether electronic portfolios designed and assessed in optimal conditions represent sufficient quality to make them useful in practice. Presumably, teachers who can reflect deeply on their work and its impact on others can improve the quality of their teaching. This study sought to answer the following research questions:

1. Does the *Rubric for Evaluating Portfolio Reflective Thinking (REPORT)* demonstrate sufficient validity and reliability for use in measuring reflective thinking in preservice teacher portfolios?
2. Do levels of reflective thinking in preservice teacher portfolios, as measured by the *REPORT*, differ between students who have and have not received instruction using a Scaffolding Intervention Tool?
3. Do elementary preservice teachers' portfolio rationale statements and reflective essays, as measured by the *REPORT*, show sufficient depth of reflective thinking to aid their growth as teachers?

Data analysis indicated that the *REPORT* instrument used in this study revealed moderate levels of interrater reliability and demonstrated sufficient content validity to be used to measure reflective thinking in preservice teacher portfolios.

Also, data indicated that members of the treatment group, who had received instruction in reflective writing, scored significantly higher on five of the six domains and

on the total score than members of the control group, who had not received instruction. There was no significant difference between groups on the Planning domain.

Analysis of the overall levels of reflection in the portfolios of both groups showed that a substantially higher percentage of preservice teachers in the treatment group (47%) wrote reflective statements that reached high levels of reflection than did the preservice teachers in the control group (6.7%). Mann-Whitney *U* comparisons supported the conclusion that preservice teachers with instructional intervention in reflective writing could demonstrate their own development in the areas of knowledge, instruction, and professional growth using more in-depth reflection than could preservice teachers who had not had this instruction. Implications for practice and further research are provided.

## Table of Contents

<b>Chapter 1: Introduction .....</b>	<b>1</b>
Overview of the Study .....	1
Background on the Problem .....	2
Theoretical Framework .....	6
Problem Statement .....	9
Purpose of the Study .....	10
Research Questions .....	12
Definitions .....	14
Chapter Summary .....	16
<b>Chapter 2: Review of the Literature .....</b>	<b>18</b>
Overview of Literature .....	18
Reflective Thinking in Teacher Education .....	18
Portfolios in Preservice Teacher Education .....	23
Portfolios as Performance Assessment .....	29
Summary of Literature .....	44
<b>Chapter 3: Methodology .....</b>	<b>45</b>
Overview of Study Methodology .....	45
Setting .....	46
Participants .....	51
Materials and Instruments .....	52
Study Procedures .....	56
Data Collection and Analysis Methods .....	58
Study Assumptions and Delimitations .....	61



Summary of Study Methodology .....	62
<b>Chapter 4: Results .....</b>	<b>64</b>
Overview of Results .....	64
Preliminary Findings on Sample and Instrumentation .....	65
Findings on Research Question 1 .....	68
Findings on Research Question 2 .....	70
Findings on Research Question 3 .....	72
Summary of Study Findings .....	74
<b>Chapter 5: Discussion .....</b>	<b>76</b>
Overview of Study Discussion .....	76
Interpretation of Findings .....	79
Implications for Practice .....	86
Study Limitations .....	89
Implications for Future Research .....	90
Summary of Study Discussion .....	94
<b>References .....</b>	<b>97</b>
<b>Appendices .....</b>	<b>111</b>
<b>Vita .....</b>	<b>119</b>

## List of Tables

<b>Table</b>		<b>Page</b>
1	Pearson $r$ Interrater Reliabilities for REPORT Ratings across Domains .....	69
2	Results of $t$ -test Comparisons of Experimental and Comparison Groups .....	71
3	Number and Percentage of Portfolios Reaching High Levels of Reflection .....	73
4	Mann-Whitney $U$ Comparison of Reflection Levels by Domain .....	74

## Chapter 1: Introduction

### Overview of the Study

Portfolios designed to measure preservice teachers' standards-based competencies, growth, and reflective ability have become ubiquitous in teacher education programs across the United States. Although proponents tout their value to promote the reflective thinking of novice teachers and imply that such thinking improves teachers' practice (Milman, 2005), few studies have confirmed these assertions by directly measuring in-depth reflection or describing conditions that develop it. Research is needed to validate effective evaluation tools that measure preservice teacher reflective capability (Yao, Thomas, et al., 2008) and to see if portfolios do, indeed, promote reflective practice. This study contributes to this literature by testing an assessment instrument to measure reflective thinking in portfolios and by examining the effects of a scaffolding intervention on the levels of reflection in undergraduate elementary preservice teachers' electronic standards-based exit portfolios.

The study employed a quasi-experimental design that compared the levels of reflective thinking in portfolios of preservice teachers who had and had not received an intervention to teach portfolio reflective writing. A control group consisting of 15 randomly-selected preservice teacher portfolios constructed without any intervention was compared to an experimental group of 15 portfolios randomly drawn from a cohort of preservice teachers who had been instructed using an intervention, the *Portfolio Reflective Writing Guide* (Appendix B). Chapter 1 outlines the purpose of the study, describes the background of the problem, explains its theoretical framework and provides the research questions that focus the research.

## **Background on the Problem**

Teacher education programs at U.S. colleges and universities are increasingly required to provide evidence that the teachers they produce demonstrate the knowledge, skills, and dispositions to ensure that all students learn at high levels (Derham & Diperna, 2007). In addition to content knowledge and pedagogical skills, budding teachers must demonstrate the ability to think carefully about the impact of their teaching on others. Organizations such as the Interstate New Teacher Assessment and Support Consortium (INTASC) outline expectations for novice teachers that include reflective thinking. Theoretically, teachers who can reflect on their practice can identify areas for improvement and make adjustments to instruction in order to increase student learning outcomes (Lyons, 1998).

Most states require teacher education graduates to pass certification exams in order to obtain licensure and teach in public schools. For the past several decades, performance assessments have been used in addition to standardized tests to measure preservice teacher quality. Preservice teacher portfolios are one vehicle frequently employed to develop and document preservice teachers' reflective capabilities. The claim is often made that portfolios demonstrate teacher candidates' growth over time and encourage reflective thinking, thus ensuring high quality teaching (Milman, 2005; Reidinger, 2006; Ring & Foti, 2006).

**Portfolios and teacher professionalism.** For the past several decades, teacher quality has been examined and often criticized (Darling-Hammond & Youngs, 2002). Strident voices continually call for education reform and higher achievement. The critical gaze of politicians and policymakers often focuses on colleges of education as a major

source of the problem, claiming they do not produce candidates who are prepared to produce high achieving students. Teacher education programs and certification rules have come under attack, and rhetoric calling for alternative certification and more rigorous teacher preparation programs has escalated in newspapers and publications (Barone & Morrell, 2007).

With each new wave of reform, teacher education leaders have intensified efforts to enhance public perception of teacher education quality and bolster a sense of professionalism. Many teacher education programs promote a vision of the teacher as “reflective practitioner.” In the 1980s and 1990s, teacher portfolios emerged as a more authentic tool to showcase the teacher’s growing development and expertise (Lyons, 1998).

The portfolio’s genesis as a means to ascertain teacher quality was the Teacher Assessment Project at Stanford University in the early 1990s (Wolf, 1991). The work done at Stanford later morphed into the National Board for the Professional Certification of Teachers (NBPCT) (Shulman, 1998; Wolf, 1991). Candidates for national board certification must create portfolios to demonstrate a range of capabilities, one of which is reflection. Core Proposition 5 involves systematic reflection, “Teachers think systematically about their practice and learn from experience” (NBPCT, 2009).

Writing in 1990, Tom Bird envisioned the possibilities of teacher portfolios to increase the status of the complex work that school teachers do, but he also cautioned that organizational and political systems to support such an undertaking must be in place. He wrote, “It remains to be seen whether, in any conditions, the schoolteacher’s portfolio can be useful for schoolteachers or for their evaluators. The idea appears plausible enough to

merit development” (p. 255). This study aims to explore one aspect of portfolio possibilities: the development of highly reflective teachers.

**Portfolios as performance assessment.** Though portfolios have proliferated, their promise to be a herald of a new professionalism has not been fulfilled (Lyons, 1998). While descriptive studies abound, empirical evidence for both the technical quality of portfolios as valid and reliable measures of teacher performance and the reflective value of portfolios is sparse (Burns & Haight, 2005; Carney, 2004; Delandshere & Arens, 2003; Herman & Winters, 1994; Reis & Villaume, 2002; Thompson, 2005; Yao, Thomas, et al., 2008). Several survey studies reported that electronic portfolios enhanced reflection and self-evaluation skills (Bartlett, 2006; Hicks, et al., 2007; Milman, 2005; Ring & Foti, 2006). However, that finding is disputed by studies that reported impoverished reflections composed mainly of descriptive statements rather than deep analysis (Bartlett, 2006; Delandshere & Arens, 2003; Ring & Foti, 2006; Sulzen, 2007).

**Portfolios as tools for reflection.** Many researchers claim that one major benefit of portfolios is that they enhance preservice teacher reflection, but limited evidence exists for such a claim (Anderson & DeMeulle, 1998; Milman, 2005; Yao, Thomas, et al., 2008). Delandshere and Arens (2003) employed case study methodology to examine the quality of the evidence in paper portfolios from three teacher education programs. They posited that portfolios lack theoretical orientation and that their organization around standards as discrete descriptions of performance represented a fragmented view of teaching.

Evidence to support the importance of reflective practice for teachers and teacher educators is plentiful in the research literature (Gordinier, Conway, & Journet, 2006;

Orland-Barak, 2005; Van Manen, 1977); however, very few studies have specifically investigated the quality and nature of the reflective statements themselves in electronic portfolios (Ring & Foti, 2006; Tillama & Smith, 2006). Specifically, studies that include valid and reliable instruments designed to measure levels of reflective thinking are rare (Orland-Barak, 2005).

One reason for the dearth of research on the quality of teacher portfolios is that they vary greatly from one institution to the next. In addition to multiple purposes for portfolios, institutions employ myriad organizational structures, content requirements, media for delivery, and assessment methods. Zeichner and Wray (2001) outline a conceptual framework that includes the following “critical dimensions of variation” found in portfolio construction: purpose, who determines content, organizational structure, nature of the interaction surrounding portfolio construction, role of cooperating teachers in feedback, audience for the portfolio once complete, and assessment methods (p. 617). They posit that it is necessary to describe the particular conditions present in any study that aims to determine the nature and quality of portfolio evidence.

Other researchers have highlighted the necessity of instruction and supervision as essential conditions for the development of reflection through preservice teacher portfolios (Amobi, 2006; Loughran & Corrigan, 1995; Riedinger, 2006). Regardless of how the portfolios are structured, student teachers must receive specific instruction regarding the reflection process, the characteristics of the reflective writing genre, and the value of reflection for future growth and improved practice (Borko, Michalec, Timmons, & Siddle, 1997; Ducharme & Ducharme, 1996; Mansvelder-Longayroux, Beijaard, Verloop, & Vermunt, 2007). Scaffolding tools such as reflective prompts, peer editing

sessions, and clearly defined rubrics may be necessary to transform portfolios from mere scrapbooks into effective tools for deep reflection.

### **Theoretical Framework**

This study is grounded in three theoretical foundations concerning reflection as it relates to effective teaching. First, it is based in theory that reflection contributes to growth in effective teaching and thus, is important to foster and measure. Second, it attends to theory on the measurable components of reflection. Finally, it is based on theory that reflection will be more likely to occur when certain conditions are in place.

**Reflection to promote effective teaching.** Beginning with Dewey's (1933) concept of reflection as rational problem solving, teacher educators have considered reflective thinking essential to improving practice. Schön's (1983, 1987) work increased the focus on reflection as a way for teachers to frame and solve problems within the complex context of teaching situations (Loughran, 2002). Within teacher education, a body of research supports the impact of effective reflection on teacher's understanding and application of "wisdom-in-practice" gained as they analyze and articulate multiple views on problems encountered in the classroom (Loughran, 2002; Spalding & Wilson, 2002). By careful reflection on experience over time, teachers develop professional knowledge and connect theory to practice (Lee, 2008; Loughran, 2002; Van Manen, 1977). In essence, effective reflection leads to effective teaching (Loughran, 2002).

**Components of reflection.** Reflection as a construct has eluded precise definition, and attempts to measure it have produced ambiguous results (Rodgers, 2002). Van Manen (1977) offered one of the first taxonomies for describing reflection. Rooted in three epistemological frameworks or interpretations of "the practical," Van Manen



proposed three levels of reflectivity: technical-rational, deliberative, and critical (Boody, 2008). Technical-rational reflectivity, grounded in empirical-analytical theory, is concerned with determining how efficiently methods and means accomplish the predetermined ends or objectives of instruction. In other words, how effectively has the teaching method achieved the goals set for it by theory or outside authority? Van Manen's (1977) second level of reflectivity (deliberative), emerging from a phenomenological-hermeneutic stance, calls for "an interpretive understanding both of the nature and quality of educational experience" (p. 226). In this level, teachers recognize their own value commitments to a particular interpretive framework as they make judgments about education practices (curriculum, methods, etc.). Finally, Van Manen proposed a higher level of reflectivity aimed at pondering "worthwhile educational ends" on the basis of "justice, equality, and freedom" (p. 227). In this critical level, teachers consider the political, moral, and ethical impact of established educational practices. Van Manen (1977) stated,

On this level, the practical addresses itself, reflectively, to the question of the worth of knowledge and to the nature of the social conditions necessary for raising the question of worthwhileness in the first place. The practical involves a constant critique of domination, of institutions, and of repressive forms of authority. (p. 227)

These three categories (technical, deliberative, and critical) emerged in many of the other reflective thinking taxonomies developed by later researchers (Hatton & Smith, 1995; Robinson & Kelley, 2007; Sparks-Langer et al., 1990; Valli, 1997; Watts & Lawson, 2009). Though each of these taxonomies offers some description of levels of

reflection, none provided a rubric for measuring reflection in a quantitative fashion. Sparks-Langer et al., (1990) provided a *Framework for Reflective Pedagogical Thinking* consisting of seven levels, each described by one phrase. For example, the highest level (7) is, “Explanation with consideration of ethical, moral, political issues” (p. 27). While such a short definition is a helpful coding scheme for qualitative document analysis, it is insufficient for practitioners such as college supervisors and cooperating teachers to use in real contexts.

Assessment research provides guidelines for creating a rubric to measure reflective thinking so that valid and reliable decisions can be made based on this tool (Brookhart & Nikto, 2008). Specifically, in this study the rubric must demonstrate content validity and construct validity to enable appropriate decisions to be made about preservice teachers’ reflective capabilities (Gay, Mills, & Airasian, 2006). Further, portfolio assessment scoring procedures must establish interrater reliability for consistency across raters (Gay, Mills, & Airasian, 2006).

**Conditions that promote reflection.** A review of the reflection research reveals that most novice teachers struggle with deeper levels of reflection (Lee, 2005). Various conditions and methods to promote critical reflection emerge from the literature. First, critical reflection seems directly connected to direct contact with teaching children in field experiences (Lee, 2005). Second, reflection seems linked to both the content and the mode of communication (Lee, 2005). Some students prefer oral explanations over written reflective statements, for example. Time is also a factor; novice teachers develop deeper reflection as they gain more experience in the classroom. Other factors affecting the development of reflective thinking identified by Lee (2005) are personal background,

structure of the dialogue and questions asked, and the context of the field placement where teaching occurs. Several researchers have pointed to coaching and specific course activities (case studies, journals, portfolios, video analysis) as methods to promote critical reflection (Lee, 2008; Rodgers, 2002; Sparks-Langer & Colton, 1991; Spalding & Wilson, 2002). Several have suggested that teacher educators provide prompts or questions to guide teachers' reflective thinking and writing (Lee, 2008; Welsch & Devlin, 2006).

With respect to portfolios, if the necessary conditions exist within the context of the teacher education program to allow them to be reflective, then the likelihood that a rubric will detect growth in reflective writing is greater (Rickards, et al., 2008). When preservice teachers clearly understand the reflective purpose for the portfolio, have sufficient guidelines for structuring it, and have been taught to write using a reflective writing "genre," then one could expect the reflective statements in their portfolio to demonstrate a greater depth of reflection (Hatton & Smith, 1995). Further, a specific tool to scaffold reflective writing that contains the definition of deep reflection, descriptions of the levels in a reflective thinking taxonomy, and models of reflective statements, may enhance the value of portfolios as reflective vehicles (Spalding & Wilson, 2002).

### **Problem Statement**

Though the body of research literature on both preservice teacher reflection and portfolio assessment clearly describes many aspects of portfolio construction, specific examples of valid and reliable tools to measure levels of reflection are conspicuously absent. Analytical tools to measure teacher reflection in general are plentiful, but few are designed specifically for portfolios (Hatton & Smith, 1995; Orland-Barak, 2005;

Rickards et al., 2008; Robinson & Kelley, 2007; Spalding & Wilson, 2002; Sparks-Langer, et al., 1990; Valli, 1997; Watts & Lawson, 2009).

Because most preservice teacher portfolios are structured around sets of teacher standards, the reflective statements they contain encompass multiple competencies, thus complicating the assessment process. Measurement researchers point out that analytic scoring rubrics for performance assessments such as portfolios must include clear descriptions of the criteria (dimension or trait) to be evaluated as well as levels of performance for each identified dimension (Brookhart & Nitko, 2008). Teacher educators would benefit from a complex, yet concise tool useful for measuring levels of reflective thinking found in preservice teacher portfolios, enabling validation of portfolios for developing reflective capabilities (Yao, Aldrich, & Foster, 2008).

Due to the lack of an appropriate instrument to measure reflective thinking in preservice teacher portfolios, it is difficult to conduct research on the impact of instruction that would enhance preservice teacher reflective capability. Specifically, the value of portfolios to promote preservice teacher reflection cannot be assessed unless reflection can be measured and taught to preservice teachers. Consequently, further research is needed to examine the claims for electronic portfolios as evidence of high quality teacher reflection (Zeichner & Wray, 2001).

### **Purpose of the Study**

This study was designed to determine whether teacher portfolios can be validly and reliably assessed, to investigate the effect of an instructional tool on increasing the level of reflective thinking in elementary preservice teachers' portfolios, and to find whether electronic portfolios designed and assessed in optimal conditions represent

sufficient quality to make them useful in practice. Presumably, teachers who can reflect deeply on their work and its impact on others can improve the quality of their teaching. In order to determine the impact of an instructional intervention, it is necessary to assess the level of reflective thinking in portfolios before and after such an intervention. Because no instrument currently existed to measure the depth of reflection in preservice teacher's electronic portfolios, this study offered a rubric to measure portfolio reflective thinking and used that rubric to measure reflective thinking after instruction had taken place.

The results of this study could affect a variety of stakeholders. First, teacher education faculty design and implement the portfolio, using it to make decisions regarding teacher candidate and program quality. Findings from this study could guide faculty as they create portfolio requirements, instruct students in how to develop reflective statements, and score the completed portfolios.

Second, teacher candidates have a high stake in portfolio work. When requirements, instruction, and assessment criteria are clear, portfolio authors may think more deeply about all aspects of their work as teachers and their ability to improve K-12 students' learning.

Third, state accrediting bodies approve the portfolios as part of each teacher education program's unit assessment plan so that the college gains status as an approved program and is able to certify teachers. In addition, programs seeking national accreditation by the National Council for Accreditation of Teacher Education (NCATE) often use the portfolio as evidence for meeting its rigorous standards for preservice teacher preparation.

Fourth, PK-12 schools employ teacher education program graduates and assume that candidates have met established competency standards. Valid and reliable means of measuring portfolio reflective thinking and subsequent instruction to increase the depth of reflection can indirectly ensure that new teachers are ready to carefully analyze their own performance.

Last, students who will receive instruction from candidates as future teachers may be able to achieve important academic outcomes if their teachers can effectively gauge the impact of their teaching on student growth. Though this study did not directly address the relationship between deep reflective thinking and K-12 student learning, the goal of reflection is to promote excellent teaching and thus increase student achievement.

### **Research Questions**

This study was designed to determine whether teacher portfolios can be validly and reliably assessed, to investigate the effect of an instructional tool on increasing the level of reflective thinking in elementary preservice teachers' portfolios, and to find whether electronic portfolios designed and assessed in optimal conditions represent sufficient quality to make them useful in practice. The research was designed to address the following issues:

1. Can a research-based instrument be designed that teacher educators can use to measure reflective thinking in practice?,
2. Can an instructional intervention designed to scaffold reflective thinking increase the levels of elementary preservice teachers' reflective thinking in the electronic portfolio rationale statements and reflective essays?, and

3. Do elementary preservice teachers' portfolio rationale statements and reflective essays show sufficient depth of reflective thinking to aid their growth as teachers?

The need for Question 1 arose from an examination of the research literature on preservice teacher portfolios. Delandshere and Arens (2003) reported that portfolios contain predominantly descriptive reflective statements rather than in-depth analyses that promote critical thinking. Zeichner and Wray (2001) cautioned that teacher educators must not simply assume that portfolios increase reflection for teacher candidates. They called for a "closer study of the nature and quality of this reflection" (p. 619). Yao and Thomas, et al., (2008) pointed out that portfolios need to be validated specifically for use as reflective tools. Such close examination in research requires valid and reliable instruments for measuring reflection. While some general checklists are available in the literature, a specific tool to determine the levels of reflective thinking in preservice teacher portfolios does not currently exist. The *Rubric for Evaluating Portfolio Reflective Thinking (REPORT)* (Appendix A) used in this study attempted to delineate various dimensions of teacher reflective capability as well as levels of quality along each dimension. Demonstrated evidence that such an instrument can discriminate between lower and higher levels of reflection will provide portfolio assessors with a means to identify excellent portfolios and study the impact of instruction in reflective writing on preservice teacher reflection.

An investigation of the research literature revealed that preservice teachers are not inclined to write in-depth reflective statements on their own, without guidance and instruction (Borko, Michalec, Timmons, & Siddle, 1997; Mansvelder,-Longayroux,

Beijaard, Verloop, & Vermunt, 2007). Question 2 derived from the need to provide specific scaffolding to preservice teachers as they create their portfolios (Gordinier, Conway, & Journet, 2006). Specifically, instruction in the reflective purpose of the portfolio, teacher coaching, prompts, informal feedback, and technical support enhance levels of reflection (Zellers & Mudrey, 2007). Question 2 in this study was designed to determine the specific effect of such scaffolding on the written reflective statements contained in the preservice teachers' developmental portfolios. A final research question addresses the issue of the inherent usefulness of portfolios to increase teachers' reflection. If reflection is being measured adequately and students are given what they need to develop their levels of reflection, then portfolios developed under these optimal conditions should show that students are, indeed, able to demonstrate the high levels of reflection that could lead to their growth as teachers.

To address these issues, this study sought to answer the following research questions:

1. Does the *Rubric for Evaluating Portfolio Reflective Thinking (REPORT)* demonstrate sufficient validity and reliability for use in measuring reflective thinking in preservice teacher portfolios?
2. Do levels of reflective thinking in preservice teacher portfolios, as measured by the *REPORT*, differ between students who have and have not received instruction using a Scaffolding Intervention Tool?
3. Do elementary preservice teachers' portfolio rationale statements and reflective essays, as measured by the *REPORT*, show sufficient depth of reflective thinking to aid their growth as teachers?



## Definitions

The following terms will be used throughout this study. A clear definition of each term as it will be used in this work is essential for the reader to understand the study:

**Portfolio.** This study used the definition provided by Lee Shulman (1998), an early innovator in teaching portfolios. He wrote, “A teaching portfolio is the structured, documentary history of a set of coached or mentored acts of teaching, substantiated by samples of student portfolios, and fully realized only through reflective writing, deliberation, and conversation” (Shulman, 1998, p. 37). Each aspect of this working definition is crucial to the success of portfolios to enable teachers to improve practice and increase student achievement.

**Electronic portfolio.** An electronic portfolio is similar to a paper portfolio except that artifacts and reflective statements are stored using electronic means. Milman (2005) lists alternate terms and explains, “A digital teaching portfolio, also referred to as a digital portfolio, electronic portfolio, e-folio, webfolio, multimedia portfolio, or electronically augmented portfolio, is similar to a traditional portfolio; however, the medium used to organize and present it is different” (p. 9). According to Milman (2005), the value of the digital version of a portfolio is its ability to combine electronic media such as videos, *PowerPoints*, and spreadsheets with more text-oriented artifacts such as papers created through word-processing programs. The portfolios that were the focus of the document analysis in this study were electronic, but that feature was not directly addressed in the measurement of reflective thinking.

**Reflection.** One of the difficulties of measuring reflection is that no single agreed-upon definition exists. A practical layperson’s sense of the word is that reflection

is thinking about the actions you have or are taking and asking if they are worthwhile. Dewey's (1933) classic definition guided this study: "Active, persistent, and careful consideration of any belief or supposed form of knowledge in the light of the grounds that support it and the further conclusions to which it tends constitutes reflection thought" (p. 9). Active attention to one's own beliefs and actions is a key component of reflective thinking and were represented in both the rubric and the scaffolding tool developed in this study.

**Scaffolding.** Deriving from cognitive information-processing theory, the term "scaffolding" in relation to education refers to the support given to learners as they engage in increasingly more difficult tasks. Reflection requires complex cognitive processes. In this study, the tool used to "scaffold" preservice teachers was designed to guide their thinking through prompts, examples, collaboration, and clear assessment guidelines (the *REPORT*).

**Preservice teacher.** College students enrolled in teacher education programs but not yet graduated, credentialed, or employed are often called "preservice" teachers. In the accreditation literature (NCATE), a synonymous term is "teacher candidate" because the individual is a candidate to become credentialed at the end of the approved teacher education program. In addition, the term "student teacher" may be used more specifically to refer to a preservice teacher (teacher candidate) who is currently participating in the final clinical practicum designated in the program of study. In this study all three terms (preservice teacher, teacher candidate, and student teacher) were used to refer to the participants because they all were teacher candidates (preservice teachers) enrolled in the student teaching semester when the portfolios were created.

## Chapter Summary

This study was designed to compare the levels of reflective thinking in portfolios of preservice teachers who had and had not received an intervention to teach portfolio reflective writing. It used a rubric to measure reflective thinking in portfolios and it determined whether the levels of reflective thinking were greater after an instructional intervention designed to enhance reflective capability in preservice teachers' portfolios.

The study sought to answer the following questions:

1. Can a research-based instrument be designed that teacher educators can use to measure reflective thinking in practice?,
2. Can an instructional intervention designed to scaffold reflective thinking increase the levels of elementary preservice teachers' reflective thinking in the electronic portfolio rationale statements and reflective essays?, and
3. Do elementary preservice teachers' portfolio rationale statements and reflective essays show sufficient depth of reflective thinking to aid their growth as teachers?

Results from this study will guide teacher educators as they help preservice teachers construct electronic portfolios. The remaining chapters provide an in-depth review of the research literature on portfolios in teacher education, explain the research methods used, provide the results and an analysis, and discuss implications of the findings for teacher educators as they seek to use portfolios to promote reflective thinking in their candidates.

## **Chapter 2: Review of the Literature**

### **Overview of Literature**

Literature from three aspects of research in teacher education framed this study: research on reflective thinking in teacher education, research on portfolios in teacher education in general, and finally research on portfolios as performance assessment tools. The first of these sections serves to document the three-part theoretical foundation of this study concerning reflection as it relates to effective teaching. First, it was based in theory that reflection contributes to growth in effective teaching and thus, is important to foster and measure. Second, it attended to theory on the measurable components of reflection. Third, it was based on theory that reflection documented in portfolios will be more likely to occur when certain conditions are in place.

A second major body of research is given here to document the role of portfolios in teacher education (Zeichner & Wray, 2001). Specifically, research has explored whether portfolios can be used as formative and summative assessment. Finally, the body of evidence on portfolios as performance assessment addresses major concerns related to a range of issues, including whether evidence exists to support the claim that they promote and increase pre-service teacher reflection. This review of the literature addresses the research on reflection in teacher education generally as well as the myriad questions about the value of portfolios in teacher education.

### **Reflective Thinking in Teacher Education**

Since the time of Plato, philosophers and educators have valued reflective thinking as part of pedagogy. Plato's concept of *phronesis* (wisdom) is contrasted with a more mechanistic concept called *techne* (technique). John Dewey's work initiated the

more contemporary focus on reflective thinking conceptualized as a version of the scientific method (Dewey, 1933). Schön's (1987) distinction between *reflection-on-action* and *reflection-in-action* shaped much of the later research on reflective thinking, specifically for those in the teaching profession. Reflection is not an end in itself, but leads to improved practice as teachers consider the impact of their work on student learning, the community, and other professionals (Zeichner & Wray, 2001).

Reflection, though widely recognized as an important skill for teachers, is not easily defined (Rodgers, 2002). Dewey's (1933) early definition often serves as a foundation for other researchers attempting to define reflection. He defined reflective thought as, "Active, persistent, and careful consideration of any belief or supposed form of knowledge in light of the grounds that support it and the further conclusions to which it tends" (p. 9). This definition highlighted the need for both careful thought and consideration of evidence when reflecting.

Other researchers included more personal experiences in their conceptions of reflection and included the element of personal growth (Orland-Barak, 2005; Valli, 1997). Many emphasized the need to consider multiple perspectives when reflecting, as well as consideration of moral and ethical implications embedded in the larger social context (Hatton & Smith, 1995; Robinson & Kelley, 2007; Sparks-Langer, Simmons, Pasch, Colton, & Starko, 1990; Valli, 1997). Most concluded that deep reflection involves the ability to think carefully about the task or experience, apply evaluation criteria of some kind, engage in self-dialogue to entertain multiple perspectives, and employ critical thinking in the context of moral, ethical, and social considerations.

**Reflection to promote effective teaching.** Beginning with Dewey's (1933) concept of reflection as rational problem solving, teacher educators have considered reflective thinking as essential to improving practice. Schön's (1983, 1987) work initiated increased focus on reflection as a way for teachers to frame and solve problems within the complex context of teaching situations (Loughran, 2002). Within teacher education, a body of research supports the impact of effective reflection on teacher's understanding and application of "wisdom-in-practice" gained as they analyze and articulate multiple views on problems encountered in the classroom (Loughran, 2002; Spalding & Wilson, 2002). By careful reflection on experience over time, teachers develop professional knowledge and connect theory to practice (Lee, 2008; Loughran, 2002; Van Manen, 1977). In essence, effective reflection leads to effective teaching (Loughran, 2002).

**Components of reflection.** Reflection is a complex construct that makes its measurement a challenge. Numerous researchers offered taxonomies for measuring reflective thinking, each with a specific focus aimed at the vehicle being used to reflect (video analysis, journal writing, etc.) (Lee, 2005; Spalding & Wilson, 2002). Several authors utilized these general taxonomies as analytical tools for qualitative exploration of preservice teacher portfolios (Orland-Barak, 2005; Zellers & Mudrey, 2007), but few have offered specific rubrics to measure reflective writing in the reflective statements that accompany artifacts contained in standards-based preservice teacher portfolios (Rickards, et al., 2008).

Van Manen (1977) constructed one of the first taxonomies for describing reflection. Rooted in three epistemological frameworks or interpretations of "the practical," Van Manen proposed three levels of reflectivity: technical-rational,

deliberative, and critical (Boody, 2008). Technical-rational reflectivity, grounded in empirical-analytical theory, is concerned with determining how efficiently methods and means accomplish the predetermined ends or objectives of instruction. In other words, how effectively has the teaching method achieved the goals set for it by theory or outside authority?

Van Manen's (1977) second level of reflectivity, emerging from a phenomenological-hermeneutic stance, called for "an interpretive understanding both of the nature and quality of educational experience" (p. 226). At this level, teachers recognize their own value commitments to a particular interpretive framework as they make judgments about education practices (curriculum, methods, etc.). Finally, Van Manen proposed a higher level of reflectivity aimed at pondering "worthwhile educational ends" on the basis of "justice, equality, and freedom" (p. 227). At this critical level, teachers consider the political, moral, and ethical impact of established educational practices, including whether they further the common good for all humans.

These three categories (technical, interpretive, and critical) emerged in many of the other reflective thinking taxonomies developed by later researchers (Hatton & Smith, 1995; Robinson & Kelley, 2007; Sparks-Langer et al., 1990; Valli, 1997; Watts & Lawson, 2009). Though each of these taxonomies offered some description of levels of reflection, none provided a rubric for measuring reflection in a quantitative fashion. Sparks-Langer et al., (1990) provided a *Framework for Reflective Pedagogical Thinking* consisting of seven levels, each described by one phrase. For example, the highest level (Level 7) was, "Explanation with consideration of ethical, moral, political issues" (p. 27). While such a short definition is a helpful coding scheme for qualitative document

analysis, it is insufficient for practitioners such as college supervisors and cooperating teachers to use in real contexts.

Assessment research provides guidelines for creating a rubric to measure reflective thinking so that valid and reliable decisions can be made based on this tool (Brookhart & Nikto, 2008). Specifically, in this study the rubric had to demonstrate content validity to enable appropriate decisions to be made about preservice teachers' reflective capabilities (Gay, Mills, & Airasian, 2006). Further, portfolio assessment scoring procedures should establish interrater reliability for consistency across raters (Gay, Mills, & Airasian, 2006).

**Conditions that promote reflection.** A review of the reflection research revealed that most novice teachers struggle with deeper levels of reflection (Lee, 2005). Various conditions and methods to promote critical reflection emerged from the literature. First, critical reflection seemed directly connected to direct contact with teaching children in field experiences (Lee, 2005). Second, reflection seemed linked to both the content and the mode of communication (Lee, 2005). Some students preferred oral explanations over written reflective statements, for example. Time was also a factor; novice teachers evolved deeper reflection as they gained more experience in the classroom. Other factors affecting the development of reflective thinking identified by Lee (2005) were personal background, structure of portfolio-related dialogue and questions asked, and the context of the field placement where teaching occurs. Several researchers have pointed to coaching and specific course activities (case studies, journals, portfolios, video analysis) as methods to promote critical reflection (Lee, 2008; Rodgers, 2002; Sparks-Langer & Colton, 1991; Spalding & Wilson, 2002). Several researchers



suggested that teacher educators provide prompts or questions to guide teachers' reflective thinking and writing (Lee, 2008; Welsch & Devlin, 2006).

With respect to portfolios, if the necessary conditions exist within the context of the teacher education program to allow them to be reflective, then the likelihood that a rubric will detect growth in reflective writing is greater (Rickards, et al., 2008). When preservice teachers clearly understand the reflective purpose for the portfolio, have sufficient guidelines for structuring it, and have been taught to write using a reflective writing "genre", then one could expect the reflective statements in their portfolio to demonstrate a greater depth of reflection (Hatton & Smith, 1995). Further, a specific tool to scaffold reflective writing that contains the definition of deep reflection, descriptions of the levels in a reflective thinking taxonomy, and models of reflective statements, may enhance the value of portfolios as reflective vehicles (Spalding & Wilson, 2002).

### **Portfolios in Preservice Teacher Education**

**History and context.** Teacher education programs at United States colleges and universities are increasingly pressured to provide evidence that the preservice teachers they produce meet high standards (Derham & Diperna, 2007). Federal legislation in the form of No Child Left Behind (NCLB) requires schools to employ "highly-qualified" teachers. More recently, the federal Race to the Top competition requires states to include measures of preservice teacher effectiveness in teacher preparation programs in order to gain approved status. Teacher candidates must demonstrate attainment of the knowledge, skills, and dispositions that lead to increased K-12 student achievement. Varied assessment tools are employed to do this including course work, student teaching

practicum observations, standardized tests such as Praxis II, state-mandated teacher assessment tests, and portfolios.

Since the 1980's, teacher educators have sought to implement less-traditional paper and pencil assessments and have searched for evaluation tools consonant with constructivist teaching frameworks (Burns & Haight, 2005; Schwartz & Rolheiser, 2001). Teacher educators have increasingly used performance assessment to embed evaluation into real world contexts and align assessment with instruction (Barton & Collins, 1993; Evans, Daniel, Mikovch, Metze, & Norman, 2006; Van Sickle, Bogan, & Kamen, 2005). Within the profession, due to the influences of accrediting and licensure organizations such as the National Board for Professional Teaching Standards (NBPTS), Interstate New Teacher Assessment and Support Consortium (INTASC), and the National Council for Accreditation of Teacher Educators (NCATE), teacher educators have implemented portfolios as an exit evaluation experience for teacher education programs (Barton & Collins, 1993; Delandshere & Arens, 2003).

Based on the concept of an artist's portfolio as a collection of best work, most early teacher portfolios were paper and often consisted of three-ring binders with the addition of videos or other "artifacts" (Barrett, 2007, p. 436). This collection of artifacts was sometimes accompanied by short reflective statements explaining the work and why it was selected (Barrett, 2007).

Electronic portfolios, which utilize the capabilities of computer-based and web-based technological tools, have emerged as the newest vehicle for documenting evidence of teacher quality (Gatlin & Jacob, 2002; Strudler & Wetzel, 2005; Wetzel & Strudler, 2005). Some have claimed they are the next "great innovation in education" (Gathercoal,

Love, & McKean, 2007, p.1). Both prior experience with paper portfolios and pressure from accrediting bodies to document teacher candidates' performance on a range of standards has motivated a new interest in the use of technology to monitor and preserve evidence (Strudler & Wetzel, 2005; Thompson, 2005).

**Portfolio definitions and description.** A preliminary review of the research literature revealed no universally-accepted definition of a teacher education portfolio. In general, a portfolio is a collection of items selected to show some quality or characteristic of the author. In the literature, some definitions represented portfolios as primarily reflective and others pointed toward documentation of evidence. Shulman (1998) has defined a teaching portfolio as “the structured, documentary history of a set of coached or mentored acts of teaching, substantiated by samples of student portfolios, and fully realized only through reflective writings, deliberation, and conversation” (p. 37). Winsor, Butt, and Reeves (1999) pointed in a different direction and defined professional portfolios as “records of goals, growth, achievement and professional attributes developed over time in professional practice in collaboration with others” (p.11). Barrett (2007) defined an electronic portfolio as “a collection of authentic and diverse evidence, drawn from a larger archive representing what a person or organization has reflected and that is designed for presentation to one or more audiences for a particular rhetorical purpose” (p. 438).

Digital portfolios, also known as an electronic portfolios, e-folios, webfolios, multimedia portfolios, or electronically augmented portfolios (Milman, 2005), enable teacher candidates to include various types of media evidence including hypertext links, audio artifacts, graphics, video , and text (Barrett, 2007). Electronic portfolios can be in

the form of word-processed document on CDs, *PowerPoint* presentations with hyperlinked items, or stored online in archives supported within institutions or available through commercially-purchased portfolio products (Chambers & Wickersham, 2007; Gatlin & Jacob, 2002; Hill, 2003).

Banister, Vannatta, and Ross (2006) conducted a survey to ascertain faculty and student preferences between three commercial programs. They reported that students encountered many challenges with ease of use, uploading artifacts, viewing and assessing artifacts, and finding a good fit for the particular institution. Survey results ranked *TaskStream* as the program with the highest overall ease of use (Banister, Vannatta, & Ross, 2006).

An additional feature of e-portfolio systems is their ability to allow faculty to aggregate data for accreditation purposes. Since many colleges use portfolios to demonstrate teacher competency against standards such as those offered by NCATE, this capability may be helpful (Banister, Vannatta, & Ross, 2006). Matching the electronic vehicle used for portfolios to the specific needs of each institution is crucial to success (Wilhelm, et al., 2006; Bannister, et al., 2006). Issues of definition and description point to conflicting paradigms within the body of portfolio research.

**Conflicting paradigms related to portfolios.** The question of what constitutes teacher knowledge, skills, and dispositions is complex and often difficult to answer. Multiple purposes for portfolios exist, and criteria for assessment of portfolio quality can focus on both process and product (Mansvelder-Lonayroux, Beijaard, Verloop, & Vermunt, 2007). Several tensions exist in the portfolio research, for both paper and electronic portfolios. One tension is between portfolios used for summative purposes

such as teacher competency and standards-based accreditation and those used for formative purposes such as reflection, professional development, growth over time, and self-identity (Barrett, 2007; Strudler & Wetzel, 2008).

This paradigm conflict surfaced in the literature on electronic portfolios, specifically (Chambers & Wickersham, 2007; Strudler & Wetzel, 2008; Zeller & Mudray, 2007). Barrett and Wilkerson (2004) explained this conflict in terms of an underlying philosophical contest between logical positivist epistemology and constructivist epistemology. A related tension pitted teacher self-identity as a learner against that of accomplished professional (Hallman, 2007). If the portfolio is to represent learning over time, early artifacts may demonstrate weakness. By contrast, if portfolios are to be the showcase of a competent teacher ready for the classroom, only the best work will be included (Bartlett, 2006; Barrett & Wilkerson, 2004).

While some research was aimed at determining the quality of the evidence that preservice teachers have met specific standards (Barrett, 2007; Ring & Foti, 2003), other research highlighted portfolios as evidence that preservice teachers can reflect on their growth and articulate developing professional identity (Hallman, 2007; Hicks, et al., 2007). This tension between process-related purposes and product-related purposes can confuse budding teachers and complicates the nature of the investigation into the value of portfolios to assess preservice teacher preparedness (Barrett, 2007; Carney, 2006; Strudler & Wetzel, 2008).

**Dimensions of conditions describing portfolios.** One reason for the dearth of research on the quality of teacher portfolios is that they vary greatly from one institution to the next. Zeichner and Wray (2001) outlined a conceptual framework that included the

following “critical dimensions of variation” found in portfolio construction: purpose, who determines content, organizational structure, nature of the interaction surrounding portfolio construction including the role of cooperating teachers in feedback, audience for the portfolio once complete, and assessment methods (p. 617). They posited that it is necessary to describe the particular conditions present in any study that aims to determine the nature and quality of portfolio evidence.

The research literature revealed a variety of purposes for professional portfolios both in K-12 education and teacher education (Klenowski, 2000). One of the most common purposes for portfolios is to provide the opportunity for students to document growth over time and reflect on learning (Barton & Collins, 1993; Hill, 2003; Lynch & Pernawarman, 2004; Woodward & Nanlohy, 2004). Portfolios offer students ownership over their own learning, embedded instruction in real world contexts, and increased collaboration (Adams, 1995; Wolf, 1991). Both paper and electronic portfolios serve as evidence of teacher competency for accreditation and licensing (Cawyer & Caldwell, 2002; Pecheone, Pigg, Chung, & Souviney, 2005). In some cases, articles addressed teacher competencies in particular subject areas such as music (Bauer & Dunn, 2003), secondary English (Hallman, 2007), physical education (Horton, 2004; Lee & Hare, 2007), social studies (McCormick, Sunal, & Sunal, 2005), and foreign language (Dhonau & McAlpine, 2005). Less often, teacher candidates were reported to utilize portfolios in job searches and colleges of teacher education analyzed portfolio results for program evaluation purposes (Barrett, 2007; Lynch & Pernawarman, 2004; Milman, 2005; Ring & Foti, 2003). Finally, some studies claimed that electronic or digital portfolios offer the

promise of enabling teachers to improve their use of technology in teaching (Milman, 2005).

### **Portfolios as Performance Assessment**

**Portfolios used for summative assessment.** The primary purpose of many portfolio programs is to provide summative assessment for accountability and accreditation. Descriptive and survey studies abound that document the experiences of K-12 teachers, college faculty, and preservice teachers with portfolio development for both paper portfolios (Barton & Collins, 1993; Meyer & Tusin, 1999; Wolf, 1991) and electronic portfolios (Strudler & Wetzel, 2005).

*Concerns related to technical quality.* Empirical evidence for the technical qualities of portfolio assessment is less abundant (Barrett, 2007; Carney, 2004; Lynch & Purnawarman, 2004; Shapley & Bush, 1999). Herman and Winters (1994) offered four categories of concern regarding portfolio assessment quality: technical quality, implementation effects, fairness, and feasibility. Carney (2006) added technological issues related to electronic portfolios to Herman and Winters' (1994) list. The following section is devoted to a range of summative assessment considerations.

*Paper portfolios.* Specific concerns related to validity and reliability of paper portfolio assessment were documented in the research (Klecker, 2000; McFarland et al., 1997; Nazier, 1997). Several authors addressed technical concerns in general articles not part of a particular research study. Klecker (2000) included reliability and validity in her discussion of the Eastern Kentucky Teacher Education Portfolio. Designed around Kentucky's New Teacher Standards, the TEP Portfolio amassed evidence that teacher candidates have met the standards. Although Klecker mentioned that a reviewer may

determine a quick check for content validity by consulting a “Teacher Portfolio at a Glance” matrix, no statistical data for content validity were provided (Klecker, 2000, p. 37).

Morgan (1999) addressed the issue indirectly by explaining how a detailed rubric enhanced portfolio quality. Her narrative article chronicled the evolution of a portfolio rubric over five years and provided examples of clear and detailed portfolio rubrics. She cited data regarding student pass rates on certification tests in Texas as evidence that candidates demonstrated competency.

Several well-designed studies illuminated the issues of portfolio assessment at the college level. Naizer (1997) investigated specific validity and reliability issues in a complex correlation study. They posed three questions related to consistency between instructor and peer raters, concurrent validity with other teaching factors, and correlation to test scores. Naizer (1997) reported that very little variance in scores was due to the raters. Specific percentages of agreement between raters, however, ranged from 64% to 92%. Although the study reported very low correlation between portfolio scores and the final examination ( $r=.22$ ), the authors concluded that portfolios can be reliably graded and that findings established sufficient concurrent validity to offer support for the use of portfolio assessment.

Good and Weaver (2003) highlighted the need for multiple data sources to match teacher evaluation with program objectives. Using both an observation tool and a portfolio as measures of teacher quality, they reported high levels of internal consistency for the observation instrument but less conclusive data from the portfolios measures. Portfolio rating scores displayed less variation than the observation scores. The authors



attributed this apparent consistency of scores to interrater reliability, although no reliability data was actually calculated. The strength of this study was that it utilized two measures of teacher quality frequently found in teacher education internships: observation instruments and portfolios.

Johnson (2006) reported high levels of reliability and construct validity when scoring online portfolios in an engineering education program. Researchers adjudicated scores by discussing discrepancies of more than two points. Interrater reliability as measured by Pearson  $r$  correlations improved over three years, and adjudicated correlations for each scoring dimension ranged from 0.80 to 0.88 by the third year. However, correlation coefficients for the first year were much lower, ranging from 0.28 to 0.64. Clearly, further experience with the scoring rubric and process increased interrater reliability over time. The study also reported high coefficients of determination and found that the construct the portfolio was measuring (technical communication) “accounted for no less than 73% of the variance within the model” (Johnson, 2006, p. 283).

In addition to the quantitative research summarized here, a few qualitative studies offered cautious support for portfolio assessment in teacher education programs while raising critical issues. Winsor, Butt, and Reeves (1999) employed case study methodology filled with rich descriptions of student portfolio development and concluded that portfolios “are one giant step forward in the pursuit of authentic and effective appraisal of student teachers’ professional judgment” (p. 30). Working from authenticity criteria for establishing trustworthiness supplied by Guba and Lincoln, Tigelaar et al. (2005) underscored the value of portfolios to offer a “rich picture of teaching reality”

while they acknowledged “serious problems with the reliability of scores generated by portfolios in terms of consistency, objectivity, and comparability” (p. 596). Reis and Villaume (2002) highlighted tensions between validity and reliability inherent in portfolio assessment and questioned whether the extensive time and work involved in portfolio construction and grading was worth the effort.

Delandshere and Arens (2003) employed case study methodology to examine the quality of the evidence in portfolios from three teacher education programs. They analyzed qualitative data in the form of interviews, analysis of portfolio documents, and observations of portfolio workshops (Delandshere & Arens, 2003). They concluded that portfolios lack theoretical orientation, and their organization around standards as discrete descriptions of performance represented a fragmented view of teaching. Regarding evidence of teacher competency found within the portfolio itself they stated,

Most entries remained unexplained, and the general descriptions provided are rarely sufficient to translate the artifacts into supporting evidence. Why the artifacts were selected, what they mean, how they relate to one another, or how they constitute the evidence claimed are rarely provided in the portfolios.

(Delandshere & Arens, 2003, p. 62)

They further questioned the role of standards in shaping a vision of teaching and definitions of teacher quality. In conclusion, reports on paper portfolios were sobering and constitute a cautionary tale to offset the plethora of portfolio programs that continue to spring up.

*Electronic portfolios.* Evans et al. (2006) expressed similar concerns regarding electronic portfolios used to assess teacher education candidates in Western Kentucky. In

a detailed descriptive article that included an extensive review of portfolio research, Evans et al. (2006) pointed to the inclusion of a “times attempted” category faculty can utilize to enhance scoring reliability and validity, but they did not conduct empirical research. Several articles chronicled the benefits of electronic portfolios and suggested methods for assessing them (including detailed rubrics), but the authors conducted no studies and offered no information regarding the validity and reliability of such instruments (Goldsby & Fazal, 2001; Gatlin & Jacob, 2002).

Very little research exists regarding the validity and reliability of electronic portfolios specifically (Carney, 2006; Derham & Diperna, 2007; Yao et al., 2008). Recently, several well-designed studies have attempted to fill this gap. Writing in 2007, Derham and Diperna (2007) stated that “no published research is currently available concerning the reliability or validity of digital portfolios” (p. 364). They posited and then tested six hypotheses designed to answer the question, “Is there evidence to support the use of digital professional portfolios for assessing the instructional competencies of preservice teachers?” (Derham & Diperna, 2007, p. 367).

Using a voluntary sample of thirty elementary and secondary education graduate-level preservice teachers enrolled in a one-semester student teaching seminar, they calculated the relationship between rubric scores on their Digital Professional Portfolio (DPP) and recognized measures of teacher quality: student teacher evaluations, grade point average (GPA), *Praxis I: Academic Skills Assessment* test scores, and *Praxis II: Principles of Learning and Teaching* test scores (Derham & Diperna, 2007).

The researchers conducted correlations using Pearson product-moment correlation coefficients and calculated reliability estimates using Cronbach’s alpha and Cohen’s

kappa coefficient. Results indicated moderate significant correlations between the DPP and *Praxis II* ( $r=.39$ ) and GPA ( $r=.34$ ), supporting two of their hypotheses. No significant correlations were found between DPP and *Student Teacher Evaluations* or between DPP and *Praxis I*. Results for hypotheses related to reliability were mixed. The authors reported evidence of internal consistency (Cronbach's alpha =.80) but not interrater reliability with "a median kappa coefficient of 0.14 across all raters" (Derham & Diperna, 2007, p. 373).

Yao and Thomas et al. (2008) conducted a large and very thorough study grounded in Messick's unified, six-faceted concept of construct validity. Participants were 128 preservice teachers in the elementary cluster of a mid-western university. The purpose of the study was "to validate whether the portfolio served the purpose of documenting teacher competencies" (p. 13). Organized around a set of state standards based on the INTASC standards, the portfolio included a detailed description of which artifacts to include. Each artifact was accompanied by a written reflection explaining why the item met the standard for which it had been selected. Portfolios were assessed at three checkpoints using a rubric provided by the university (Yao et al., 2008).

Yao and Thomas, et al. (2008) included a copy of the rubric, which was created by faculty members, but did not include validity and reliability information regarding this instrument. Key to their findings was their note that the portfolio score was primarily based on the quality of the reflections, rather than both artifact quality and reflections. In addition to artifacts and reflective statements, the rubric rated formatting features such as aesthetics and writing mechanics as well as several overarching summaries they called metareflections.

Yao and Thomas et al. (2008) found support for content validity due to the fact that experts created the process to align with state standards and required portfolio items did match the standards. They found limited support for substantive validity because they determined that the cognitive processes required in the metareflections matched those required in the teacher competencies. Limited support for structural validity was found, and the authors concluded that the portfolio scores were “primarily a reflection of the preservice teacher’s reflective skills” (p. 19). The researchers also reported limited support for external validity due to weak correlations between external measures of teacher quality such as evaluation of student teachers by supervisors, *Praxis II*, GPA, ACT, and a state-administered basic skills test similar to *Praxis I*. These findings align with those found by Derham and Diperna (2007).

Overall, Yao and Thomas et al., (2008) concluded that portfolio artifacts would need to be included in order to validly measure overall preservice teacher competencies. However, they pointed out that the time required for professors to grade the artifacts, problems with reliability in grading comprehensive portfolios, underrepresentation of competencies by only one artifact for each domain, and the extensive time demands reported by students to complete the portfolio all posed barriers to use of electronic portfolios as valid and reliable measures of complete preservice teacher competencies.

Sulzan and Young (2007) created a rubric and validated a fast and reliable method of assessing preservice teacher portfolios. Due to concerns regarding excessive time requirements, this study utilized a dichotomous scoring procedure, along with one holistic subjective rating. Four raters scored seventy-five portfolios. The rubric included seventeen criteria, representing “developing portfolios exhibiting a wide breadth, but

limited depth of content” (Sulzan & Young, 2007, p. 5). The researchers found a high level of interrater reliability (0.85) with the average time required to score each portfolio being 15 to 20 minutes. This is fast and increases the likelihood that reliable assessment will realistically occur. The reader is left to wonder how such portfolios could be used for comprehensive assessment of standards-based teacher capability, since only surface level competencies are measured.

Finally, Sulzen (2007) conducted a three-study dissertation to examine the judgments faculty members made about student teaching capability while evaluating electronic portfolios. The first study identified 12 areas of expertise similar to INTASC standards which they called *The Taxonomy of Classroom Teaching*. Using a think-aloud protocol and document analysis of the portfolio rubrics, the second study concluded that faculty members made strong judgments in the areas of “content knowledge, general pedagogic knowledge, instructional design, instructional delivery, assessment, and reflection” (p. 115). Sulzen (2007) also pointed out the inconsistency of rubrics. Sulzen’s (2007) third study identified “video of teaching, lesson plans, lesson plan reflections, daily journal reflections, and a reflection on the context of teaching” as the most effective portfolio artifacts for making judgments (p. 203). He found that faculty members seldom looked deeply at course assignments already graded by other professors and recommended that most portfolio artifacts come from work completed during student teaching.

In sum, technical challenges to using both paper and electronic portfolios for summative teacher evaluation abound, whether one investigates using quantitative or qualitative frameworks. Technical quality is related to purpose, and the evidence supports

use of portfolios for some aspects of teacher performance, but not as a sole comprehensive summative tool.

***Concerns related to implementation effects.*** Carney (2006) reported on two studies conducted in PK-12 settings that link in-service teacher portfolios to student achievement. Very few studies investigated how *preservice* teacher portfolios might serve as evidence of impact on PK-12 student learning (Levitt & Schreiber, 2008). Using student teacher reflective statements, Levitt and Schreiber concluded that student teachers collected a variety of types of evidence for student learning across all subject areas. The authors reported that teacher candidates do not generally link the data they provide with claims and inferences made regarding student achievement. Further, this study gathered no external data regarding student achievement.

Impact on K-12 student learning seems to be the gold standard called for by state and national accrediting bodies (Darling-Hammond & Youngs, 2002). No studies were found that directly connected electronic portfolios with student achievement measures. Given the short-term nature of most student teaching experiences, this seems likely to be an insurmountable obstacle to demonstrating a causal relationship between student teacher portfolios and K-12 student achievement.

***Concerns related to fairness and legality.*** Increased teacher accountability raises the stakes for all teacher evaluation measures, including electronic portfolios used for teacher licensing and accreditation (Carney, 2006). Fairness considerations involve the question of who owns the work, concerns about bias, and negative impacts on diverse groups of students or teachers (Carney, 2006; Wilkerson & Lang, 2003). Wilkerson and Lang (2003) cautioned that electronic portfolios are not a safe vehicle for summative

certification decisions unless their contents and evaluation processes are carefully controlled. They also pointed out that all high-stakes assessments such as tests must demonstrate rigorous levels of validity, reliability, fairness, and absence of bias or they are vulnerable to legal challenges. Until comprehensive electronic portfolios meet these fairness-related technical and legal concerns, they may not be suitable for large-scale assessment (Wilkerson & Lang, 2003; Yao et al., 2008).

*Concerns related to technology use.* Carney's study (as cited in Sulzen, 2007) compared paper and electronic portfolios and found the mechanical challenges related to multi-media capabilities of electronic portfolios to be more severe than mechanical challenges arising in paper portfolios. Sulzen (2007) found that the commercial online template provided in his study lessened this problem, but concluded that increased media possibilities also increase the types of mechanical difficulties confronting students.

Several studies included some aspect of comparison (Bartlett, 2006; Ledoux & Henry, 2006; Woodward & Nanholy, 2004). Pecheone et al. (2005) surveyed supervisors and student teachers regarding factors that led to valuing the portfolio. Teacher candidates felt that the electronic portfolios took more time to complete than paper. These authors identified several common threads in the survey data, both positive and negative. Benefits included easy access to Web-based data, helpful and timely feedback, and a single online storage space. Time-consuming technical difficulties posed challenges to both preservice teachers and faculty members and had to be resolved before any electronic benefits could be fully realized.

Bartlett (2006) surveyed faculty members and students and reported their perceptions that creating electronic portfolios was "worth it" (p. 331). In fact, in this



study preservice teachers reported that “learning about technology was the main benefit of ePortfolios”; however, they were not sure they would incorporate portfolios into PK-12 student assessment, nor did they articulate how they would apply new technology learning to classroom instruction (Bartlett, 2006, p. 331).

Finally, Milman (2005) conducted interviews and observations with teacher education faculty and students. Her study cited several benefits as well as challenges to creating digital portfolios. Advantages included availability to anyone on the Internet, the ability to create links to other web sites as evidence for competency, and mastery of technology skills needed to create the portfolio such as uploading artifacts and writing HTML. Technology-related advantages also served as challenges. Time required to learn the technology skills, time consumed in actually constructing the portfolio, and lack of technology training served as barriers to success. Overall, several studies concluded that the benefits of electronic portfolios outweighed the negative aspects, provided sufficient resources were available to overcome the challenges (Bartlett, 2006; Chambers & Wickersham, 2007; Milman, 2005).

*Concerns related to feasibility.* Numerous articles describing various portfolios in all settings mentioned that they were time consuming for both students and teachers. Herman and Winters’ (1994) concern that time demands might compromise feasibility is echoed in recent research on electronic portfolios used to measure preservice teachers, in particular. Even if technical, legal, and technology-related issues are resolved, are the benefits worth the time and resource costs? Many of the studies already cited in this paper implied that the benefits are worth the costs, even though research is still in its infancy (Ring & Foti, 2006).

Strudler and Wetzel (2005) conducted a large survey of six institutions whose portfolio systems were mature. Using Rogers' (2003) notion of the diffusion of innovations, Strudler and Wetzel described how electronic portfolios were initiated and implemented in these six universities. Subsequent studies based on the same data set offered helpful suggestions for implementation (Wetzel & Strudler, 2005), presented student perceptions of electronic portfolios (Wetzel & Strudler, 2006), and chronicled faculty perceptions (Strudler & Wetzel, 2008). Student voices and faculty voices were heard throughout the research literature.

*Student voices.* First, students wanted to understand clearly the purpose for the portfolio (Wetzel & Strudler, 2008; Yao, Aldrich, & Foster, 2008). Regardless of the purpose, students put more effort into the work when they understood why they are expected to do it (Wetzel & Strudler, 2006; Bartlett, 2006). Second, students wanted clearly-delineated requirements and technology that worked. When either of these procedural items was weak, students became frustrated (Chambers & Wickersham, 2007; Pecheone et al., 2005; Wetzel & Strudler, 2006). Third, students appreciated the value of reflection but found it difficult and resisted doing a great deal of it (Ring & Foti, 2006; Strudler & Wetzel). Finally, students invested large amounts of time in constructing portfolios and valued feedback from professors, but also from peers and future employers (Bartlett, 2006; Chambers & Wickersham, 2007; Pecheone et al., 2005; Wetzel & Strudler, 2006). Unfortunately, students reported that professors did not give enough feedback (Bartlett) and principals did not look at portfolios in today's test-driven scene (Yao, Aldrich, & Foster, 2008).

*Faculty voices.* On the positive side, faculty members perceived that students learned through the reflection afforded in electronic portfolios (Penny & Kinslow, 2006; Ring & Foti, 2006; Strudler & Wetzel, 2008), electronic portfolios increased understanding of state and national standards (Ring & Foti, 2006; Strudler & Wetzel, 2008), and electronic portfolios made accessing and evaluating student work easier, if no technical difficulties hindered the work (Strudler & Wetzel, 2008; Wilhelm et al., 2006).

On the negative side, intense time and training were needed to implement a new assessment innovation, particularly when new technologies must be mastered (Bartlett, 2006; Penny & Kinslow, 2006; Strudler & Wetzel, 2008). While accreditation accountability is often cited by faculty as a motivator for electronic portfolio adoption, some faculty members resisted such pressure due to concerns about academic freedom and philosophical disagreements over standards-based reform (Strudler & Wetzel, 2008; Delandshere & Arens, 2003). Systematic portfolio assessment systems require intense alignment of standards, objectives, and assessment tools in syllabi, coursework, and student teaching experiences. If professor buy-in is low, it is difficult to accomplish these cooperative tasks (Penny & Kinslow, 2006).

In general, no final answer exists about whether electronic portfolios are “worth it” (Strudler & Wetzel, 2008). A need remains to carefully evaluate the efficacy of electronic portfolios based on a cost-benefit analysis in each particular context (Sulzen, 2007; Yao et al, 2008).

**Portfolios used as formative assessment.** Undoubtedly, many of the concerns related to summative portfolio assessment also apply to their use as formative assessment

tools as well. The research literature offered additional insights into portfolios designed to measure preservice teacher professional development and reflective thinking.

*Portfolios as tools for reflection.* Several researchers have investigated the ways in which paper portfolios scaffold reflection and promote professional development (Borko, Michalec, Timmons, & Siddle, 1997; Wenzlaff & Cummings, 1996). Hartmann (2004) utilized case study methodology to describe how paper portfolios enabled one professor to increase reflective thinking in prospective secondary math teachers. He identified specific ways in which student teacher beliefs affected actual teaching practice. One of the few studies that clearly connects preservice teachers' portfolios to K-12 student learning, this study narrated one college student's experience with connecting theory to practice in the field of math (Hartmann, 2004).

While many of the assumptions and research findings garnered from paper portfolios may apply to electronic portfolios, research related specifically to the capacity of electronic portfolios to measure and enhance preservice teacher reflection is not abundant. Several survey studies reported that electronic portfolios enhanced reflection and self-evaluation skills (Bartlett, 2006; Hicks et al., 2007; Milman, 2005; Ring & Foti, 2006). However, that finding is balanced by studies that reported impoverished reflections composed of mainly descriptive statements rather than deep analysis (Bartlett, 2006; Ring & Foti, 2006; Sulzen, 2007). Yao, Aldrich, and Foster (2008) suggested that electronic portfolio reflections are richer when based on actual field experiences.

Rickards and his colleagues at Alverno College (2008) narrated the story of their experience with electronic portfolios at their institution. Their action research study offered useful insights for teacher educators designing reflective portfolios. Detailed

descriptions of the context, reflective prompts and protocols, and a rating scale designed for use in that particular context illustrated possible methods for developing preservice teacher reflective thinking that might be applied to other contexts.

*Assessment of reflective thinking in preservice teacher portfolios.* Overall, both descriptive and qualitative studies addressed the particular problem of whether portfolios are useful tools to produce deep reflection in teacher educators. While some concluded that they hold promise, few offered substantial empirical evidence (Zeichner & Wray, 2001).

Many portfolio studies are grounded in a constructivist epistemological framework that makes assessment processes particularly difficult. Pamela Moss (1998) proposed a hermeneutic, interpretive framework for evaluating portfolio evidence. This approach seemed particularly effective for complex cognitive constructs such as reflective thinking. On the other hand, particular assessment procedures that emerged from this approach required extensive time commitments on the part of both portfolio creators and assessors. Carney (2007) called for rigorous methodology in studies of portfolio assessment, whether measured using psychometric concepts or those embracing a hermeneutic framework.

Construct validity is difficult to establish for complex constructs such as attitudes and cognitive processes such as reflection. Clear descriptions of both individual domain criteria and levels of performance quality for each criterion are crucial to creating scoring rubrics for performance assessments (Brookhart & Nitko, 2008; Popham, 2006). Popham (2006) stated, "It is difficult to devise rubrics that embody just the right level of specificity" (p. 248). Popham echoed the concerns of others when he pointed out that in

addition to technical challenges; portfolio evaluation requires large amounts of time (Popham, 2006).

### **Summary of Literature**

Overall, research indicated that portfolios can enhance reflection for preservice teachers and contributed to their professional development. They offered challenges as well, including conflicting purposes, technical difficulties, and assessment complexities. In addition, novice teachers such as teacher candidates often found deep reflection difficult, and analysis of most portfolio reflective statements revealed shallow descriptive explanations. Teacher candidates often perceived that the portfolio's main value was to enable them to gain employment, a conclusion not supported by research. In addition, portfolio creators lacked specific guidelines as they constructed their portfolios. They also reported frustration with the lack of clear assessment guidelines and tools, such as valid and reliable rubrics. Portfolio creation requires an enormous time investment and the benefits must outweigh the challenges for all stakeholders if they are to remain a viable method to assess teacher quality and reflective capability.

## Chapter 3: Methodology

### Overview of Study Methodology

This study was designed to compare the levels of reflective thinking in the electronic portfolios of preservice teachers who have and have not received an intervention to teach portfolio reflective thinking and writing. Chapter 1 provided an overview of the entire study, including the background of the problem, its theoretical framework, the purpose of the study, problem statement, research questions, and definitions. Chapter 2 summarized the extant literature on reflective thinking and portfolio assessment, including how the two are related.

The purpose of this chapter is to describe the design, methodology, and data collection and analysis of a study designed to examine the effects of a scaffolding intervention on the levels of reflection in undergraduate elementary preservice teachers' electronic, standards-based, exit portfolios. As described in Chapter 1, this study sought to answer the following research questions:

1. Does the *Rubric for Evaluating Portfolio Reflective Thinking (REPORT)* demonstrate sufficient validity and reliability for use in measuring reflective thinking in preservice teacher portfolios?
2. Do levels of reflective thinking in preservice teacher portfolios, as measured by the *REPORT*, differ between students who have and have not received instruction using a Scaffolding Intervention Tool?
3. Do elementary preservice teachers' portfolio rationale statements and reflective essays, as measured by the *REPORT*, show sufficient depth of reflective thinking to aid their growth as teachers?

This chapter describes the setting and participants, the specific research design, the instruments used to answer the research questions, and the data collection and analysis procedures. It further highlights research assumptions and limitations that guide understanding of the generalizability of this study to different settings.

## **Setting**

**General context.** The context of this study was the elementary teacher education program at a small, liberal arts college located in the southeast. The college is located in a rural community of approximately 4,000 residents (1,938 males and 2,100 female) (<http://factfinder.census.gov>). The community is predominantly white (97.9%). Total college enrollment in all undergraduate programs is 965, with 44% male and 56% female. Education is one of the top six majors at the college.

The college offers two degrees in early childhood (Bachelor of Arts, or BA, in Elementary Education, P-5; and Bachelor of Science, or BS, in Early Childhood) that lead to initial teacher certification in the state in which the college is located. The BA program is housed in the Teacher Education Program of the college and serves traditional residential students enrolled in a 4-year liberal arts program. The BS program is also housed in the Teacher Education Program of the college but can be characterized as a degree-completion program. Entrance requirements include 58 semester hours of credit in content area courses. The BS program provides 62 hours of professional and methods courses. Both Early Childhood Education (ECE) programs end with 15 credits of student teaching. Participants in the study were senior student teachers from both the BS and the BA early childhood programs enrolled in the second of two full-time clinical practice experiences.



**Program description.** Participants in both programs undertake a variety of methods courses and field work experiences that culminate in the student-teaching experience. One requirement for successful completion of the student-teaching semester is to construct an electronic portfolio (Teacher Education Program Developmental Portfolio) that includes various artifacts created throughout the teacher education program, rationale statements explaining why these artifacts constitute evidence that teachers have met standards, and reflective essays highlighting preservice teachers' learning, connections of theory to practice, and ability to identify areas for improvement in future practice.

The portfolio is organized around 12 Teacher Standards designed by the Teacher Education Program (TEP) to describe the knowledge, skills, and dispositions expected of its program graduates. These 12 standards are organized into six domains: Knowledge, Planning, Instruction, Assessment, Classroom Environment, and Professional Development. For each domain, student teachers must include both required and self-selected artifacts to offer evidence that they have met the standards in that domain. In addition, they complete reflective essays at various points in the program. The reflective essays ask the preservice teachers to connect theory to practice, analyze and describe their own strengths and weaknesses, and apply understandings from the college's Biblical frame of reference to the teaching experiences represented in the artifacts (*Student Teaching Handbook*, Fall 2009, p. 46). Portfolios are graded as either Pass or Fail based on a designated rubric.

As noted in Chapters 1 and 2, preservice teacher portfolios such as the one used in this program are ubiquitous in U.S. teacher education programs. The claim is often made

that portfolios demonstrate teacher candidates' growth over time and encourage reflective thinking, thus ensuring high quality teaching (Anderson & DeMeulle, 1998; Milman, 2005; Riedinger, 2006; Ring & Foti, 2006); however, the quality of the reflection has been questioned in the literature (Delandshere & Arens, 2003; Zeichner & Wray, 2001).

One reason for this phenomenon is that portfolios exist in myriad forms and for multiple purposes. Zeichner and Wray (2001) outline a conceptual framework that includes the following "critical dimensions of variation" (p. 617) found in portfolio construction: purpose, who determines content, organizational structure, nature of the interaction surrounding portfolio construction, role of cooperating teachers in feedback, audience for the portfolio once complete, and assessment methods. They suggest that it is necessary to describe the particular conditions present in any study that aims to determine the nature and quality of portfolio evidence. Therefore, the following description of the TEP Developmental Portfolio process and requirements is essential to understanding the study's purpose and design.

Zeichner and Wray's (2001) first condition is purpose. In the TEP where this study takes place, the portfolio serves a dual purpose: both reflection and evidence for competency. Its primary aim is to offer preservice teachers the opportunity to reflect on their development throughout the program; however, preservice teachers are asked to provide an explanation for how the particular artifact they have included constitutes evidence for the performance standard under which it is located (*Student Teaching Handbook, Fall 2009*, p. 41). Throughout the portfolio coursework and training, instructors emphasize that the purpose of the portfolio is largely one of formative

assessment and careful reflection is the goal. Inevitably, however, other purposes such as use as a resource to gain employment enter the discussion.

The second condition offered in Zeichner and Wray's (2001) framework involves content selection, namely, *who* selects the required artifacts. In this TEP, preservice teachers receive a list of required portfolio artifacts, but they are also asked to submit additional self-selected artifacts. Required artifacts are designated for each domain but student teachers may add additional evidence under each domain, as desired.

Closely tied to the issue of who selects content is Zeichner and Wray's (2001) third dimension: organizational structure. The TEP Developmental Portfolio under examination in this study is structured around the department's 12 Teacher Standards, organized into six Domains, as described earlier. Such a standards-based format is common in teacher education programs due to the influences of accrediting requirements (Barrett & Wilkerson, 2004). Paper formats and electronic formats are both utilized and the TEP Developmental Portfolio in this study is constructed using a commercially-available Web-based portfolio system housed in an online software system called *LiveText*. *LiveText* offers a range of services to colleges of education, including assessment data management, portfolios, and report generation for accreditation. Teacher education program candidates purchase *LiveText* during orientation activities when they enter the program and receive further instruction for using this technological tool in the technology course and from professors of methods courses throughout the program.

A fourth aspect or condition of portfolio construction identified by Zeichner and Wray (2001) is the social interaction experienced by portfolio authors (teacher candidates) throughout the portfolio development process. The amount of peer mentoring

and discussion, professor feedback, technology support, and involvement of outside mentors such as cooperating teachers can contribute to, or detract from, the quality of reflection in the portfolios (Carney, 2007; Rickards, et al., 2008; Wray, 2007). Currently, most of the interaction surrounding the portfolio takes place in a course entitled *EDU 480 Senior Integration Project*. The portfolio and a problem-based research paper are the two main course assignments, and teacher candidates read a required text, discuss the portfolio rubric, investigate sample portfolio statements, and ask advice from the course instructors about how to select artifacts, attach them in *LiveText*, and write the reflective statements. Further, they may submit drafts of the portfolio to the professors at any time during portfolio development before the final due date. In addition, they may submit the portfolio to their college supervisor during the student teaching experience for additional feedback. No formal procedures exist for this interaction, except for those mentioned above. Often, teacher candidates will ask each other for help or discuss requirements, but this is not a structured part of the portfolio process in this particular TEP program. The cooperating teachers are not involved in either discussion or evaluation of the TEP Developmental Portfolio.

It is this fourth dimension of portfolio development that served as the nexus for the intervention in this study. An intervention tool, the *Portfolio Reflective Writing Guide* (see Appendix B) describes the levels of deep reflection outlined in the research, offers prompts and questions to scaffold reflective writing, and provides samples of various levels of performance. It was hypothesized in this study that if teacher candidates understand the purpose, content, structure, and evaluation methods of the portfolio, this scaffolding guide will lead them to write at higher levels of reflection. Study procedures

outlined below offer further description of how the *Portfolio Reflective Writing Guide* was used.

Zeichner and Wray (2001) suggest two final and related dimensions of portfolio design: presentation and assessment. The portfolio is often presented to peers and supervisors at the conclusion of student teaching and may be assessed at this time or assessed later. Numerous variations for both portfolio presentation and assessment are possible. The portfolio in this study is presented by the preservice teacher to the college supervisor first electronically and then individually in a conference called a “Connect-the-Dots” conference. During this conference, the preservice teacher and the college supervisor discuss the reflective statements and essays in order to summarize strengths and weaknesses. The college supervisor evaluates the portfolio using a rubric on a pass/fail basis. It is not the same rubric as the *REPORT* developed for this study.

Taken together, these “dimensions of variation” create a context within which the portfolio is developed and serve as multiple variables affecting the quality of the portfolio reflective writing. While it was impossible to control for all extraneous variables, the preceding descriptions serve to clarify this study’s scope.

### **Participants**

Participants in this study were preservice teachers enrolled in the final student teaching semester in two approved teacher education programs (BA and BS) that lead to initial certification in Elementary Education, Early Childhood (P-5) in a southeastern state. For this study, 15 participants for the control group were randomly selected from the population of graduates who completed their program between May 2007 and December 2009, and whose portfolios are available in the *LiveText* archive. The

treatment group consisted of 15 participants randomly selected from the preservice teachers enrolled in their final student teaching semester during the spring semester of 2010. A table of random numbers was used to select the two groups for the study. The total number of participants for the study was 30.

### **Materials and Instruments**

In order to answer research question 1, an instrument called the *Rubric for Evaluating Portfolio Reflective Thinking (REPORT)* was developed to measure the levels of teacher reflective writing in both the rationale statements and the reflective essays. The *REPORT* (see Appendix A) contains three categories of reflective thinking drawn from the research literature.

Developing an instrument to measure preservice teacher reflection is a challenge that few others have undertaken. Several researchers point to the fact that reflection is very difficult to define (Hatton & Smith, 1995; Rodgers, 2002). Further, since precise definitions are scarce, it is difficult to operationalize conceptions of reflection in assessment instruments. Hatton and Smith (1995) summarized this challenge clearly when they stated,

For a start, **definitions of reflection**, especially of the critical form, are often inappropriate or inadequate, and it is clear that the terms are extremely difficult to render operational in questionnaires and other research instruments. Then it would appear that it has been a considerable challenge to develop **means for gathering and analysing data** so that the evidence shows unequivocally that reflection has taken place. (p. 39)

The instrument developed for this study, the *Rubric for Evaluating Portfolio Reflective Thinking (REPORT)* combines concepts from several researchers. First, Van Manen (1977) proposed three “levels of reflectivity of deliberative rationality”: technical, practical, and critical (p. 226). Technical reflection examines whether means and methods (instructional procedures, for example) lead to predetermined ends (curriculum objectives), but those ends are not examined. The second level, practical reflection, examines both means (methods) and ends (objectives) and uses pragmatic criteria based on the performance outcomes. In other words, did instructional procedures and curriculum objectives lead to enhanced achievement or performance? Finally, Van Manen’s (1977) third level, critical reflection, examines the moral, ethical impact of educational practices and goals. This level of reflection examines both means and ends to determine their worth in view of the values of justice, equality, and freedom.

Later, Hatton and Smith (1995) proposed five levels of reflection built on Van Manen’s (1977) three categories and Schön’s (1983) two notions of “reflection-in-action” and “reflection-on-action”: technical, descriptive, dialogic, critical, and contextualization of multiple viewpoints (Hatton & Smith, 1995, p. 45). Valli (1997) included ideas from Van Manen (1977) and Schön (1995), and, similar to Hatton and Smith (1995), offered five types of reflective thinking in teacher preparation programs: technical reflection, reflection-in-action and on-action, deliberative reflection, personalistic reflection, and critical reflection. Valli’s (1997) addition of personalistic reflection asks the individual to attend to inner growth as well as outward impact on both technical competence, recognition of multiple perspectives, and examination of worth based on an interpretive framework that includes value commitments from a moral and ethical perspective.

Finally, Sparks-Langer et al. (1990) offered a simple hierarchical scale called the *Framework for Reflective Thinking*. This scale consists of seven levels, each described by a single phrase. The lowest level (Level 1) states, “No descriptive language” (Sparks-Langer et al., p. 55). The highest level (Level 7) states, “Explanation with consideration of ethical, moral, political issues” and reflects the same concept of critical reflection found in multiple studies on reflective thinking (Sparks-Langer, et al., p. 55). None of these taxonomies for reflective thinking were created specifically for portfolio assessment, and none seemed uniquely suited to assess standards-based preservice teacher reflective portfolios.

The researcher-created instrument used in this study encompasses Van Manen’s (1977) three levels, Hatton and Smith’s (1995) inclusion of the dialogic (multiple explanations for actions), and Valli’s (1997) focus on personal growth. It also includes a level of critical reflection that asks preservice teachers to consider the larger social context and the moral and ethical impact of the expectations of their own profession (Nagle, 2009). Both Sparks-Langer et al. (1990) and Hatton and Smith (1995) pointed out that reflection cannot be conceptualized as a simple continuum, so a hierarchical scale is not sufficient to measure reflection. To that end, the *REPORT* contains three types of reflection (technical/descriptive, personal growth, and dialogic/critical), each with three levels of depth. Scores on three types and levels were calculated for each content domain in the portfolio (knowledge, planning, instruction, assessment, classroom environment, and professional development).

Scoring procedures for the *REPORT* are holistic, as is recommended for portfolios in the research literature (Johnson, Mims-Cox, & Doyle-Nichols, 2006; Meeus, Petegem,



& Engels, 2009). Though the *REPORT* consists of six domains, each with three types of reflection and three levels, raters were instructed to read each domain in the portfolio as a whole and mark one score for each type of reflective thinking holistically. Raters scored each type of reflection (technical/descriptive, personal growth, and dialogic/critical) on a scale ranging from 0-3. Category scores were added within each domain to obtain a domain score. Then the scores on individual domains were summed to calculate a total portfolio score (see Appendix B). While this may seem more like analytic scoring than holistic (Popham, 2006), the stance the raters took toward the reflective statements and essays encompassed the body of work as a whole within each domain to arrive at the scores.

Construct validity is difficult to establish for complex constructs such as attitudes and cognitive processes such as reflection. Clear descriptions of both individual domain criteria and levels of performance quality for each criterion are crucial to creating scoring rubrics for performance assessments (Brookhart & Nitko, 2008; Popham, 2006). Popham (2006) stated, “It is difficult to devise rubrics that embody just the right level of specificity” (p. 248). The *REPORT* was designed to be psychometrically sound (Carney, 2007) and to mitigate concerns faculty expressed regarding ease of use for assessment (Strudler & Wetzell, 2008; Sulzen, 2007).

Content validity for the *REPORT* was demonstrated through expert analysis and verification. An early draft of the *REPORT* was sent in the fall of 2009 to eight experts recognized for their expertise in portfolio assessment in teacher education through research published in peer-reviewed articles. Each expert was asked to evaluate the content of the rubric as well as the descriptions of levels of performance, sample

reflective statements, and a scoring guide. Revisions were made on the basis of expert comments. In order to complete preliminary interrater reliability calculations, two raters each received training on how to use and score the *REPORT* and scored 10 portfolios drawn from the portfolio archives stored in *LiveText*. Interrater reliability was computed using a Pearson  $r$  correlation (Gay, Mills, & Airasian, 2006). Discrepancies were discussed with a goal of achieving 80% or greater interrater agreement.

### **Study Procedures**

The design for this study is a variation of a quasi-experimental design known as the Cohort Design (King & Roblyer, 1984). This design is intended for use when it would be unethical to withhold treatment (instruction) from one group in order to form a control group as in a traditional experimental design. King and Roblyer (1984) state, “The cohort design compares the performance of students in one instruction group with their cohorts who did not receive the instruction” (p. 26). In King and Roblyer’s (1984) Cohort Design, cohort groups are students in the same grade in each of two years. For example, two groups are composed of students in first grade one year at a specific school and then students in the first grade the next year, both of whom received the same outcome measure at the time. The design in this study differs from King and Roblyer’s design in that the cohorts are two separate groups randomly drawn from a pool of participants, one group who completed the portfolios before the treatment and another group who completed it after the treatment. The same instrument is used to measure performance, but the measurement is after the fact.

In this study, the two cohorts are groups of elementary preservice teachers, one who constructed a portfolio without instruction regarding reflective writing and the other

who had the instruction (treatment). Both groups (n=15) were randomly selected using a table of random numbers. The control group was randomly selected from the group of preservice teachers who completed their portfolios prior to the Spring 2010 semester and whose portfolios are entered into an archive of Developmental Portfolios housed in *LiveText*. The treatment group was a cohort of 15 preservice teachers randomly selected from the group of student teachers who had received the instructional intervention during enrollment in two sections of the Spring 2010 course, *EDU 490 Student Teaching Seminar*. Since it was possible that the groups differed in some way, GPA was used as a pretest and the two groups were compared using a *t* test (Gay, Mills, & Airasian, 2006). Letters of consent to participate in the study were obtained from all 30 participants.

The treatment consisted of an instructional intervention, the *Portfolio Reflective Writing Guide*, designed to assist preservice teachers with writing reflective responses to their own work (see Appendix B). The treatment group received a single, 1-hour instructional session composed of the following activities: (a) a short introduction using the *Portfolio Reflective Writing Guide* (see Appendix B), (b) an explanation of different types and levels of reflection based on research using the *REPORT*, (c) a list of prompts and questions designed to promote higher levels of reflection, and (d) discussion with a partner of draft reflective statements. The intervention took place during two, 1-hour sessions of *EDU 490 Student Teaching Seminar* conducted in March of 2010, one for each section of the student teaching seminar course. Since there are two sections of the course, one rater trained in the *REPORT* observed each treatment session to ensure implementation fidelity.

To ensure ethical treatment of human subjects, this study obtained approval from the university's institutional review board, all data were kept confidential, and research results and conclusions do not include information on individual participants, so no pseudonyms had to be used. Each participant signed an Informed Consent Form prior to data collection and these forms were kept secure at the research site. Scores from the *REPORT* were not included in student portfolio grades and the raters did not score the portfolios for the study until after they had been graded using the regular scoring rubric provided in the course syllabus. Candidates had completed the final student experience before the portfolio study commenced to reduce any perceived pressure that participation would affect portfolio or course grades.

### **Data Collection and Analysis Methods**

In order to answer research question 1, the researcher-developed instrument, *Rubric for Evaluating Portfolio Reflective Thinking (REPORT)*, was used to rate the portfolios of the control group and the treatment group. The researcher obtained the portfolios from the archive of portfolios housed in *LiveText*, the commercially-purchased assessment system adopted by the college teacher education program. The portfolios were submitted to the researcher by each participant using his or her *LiveText* accounts. The portfolios were printed to facilitate ease of scoring and printed *REPORT* score sheets were provided for each portfolio. Each rater scored all 30 portfolios at the conclusion of the spring 2010 semester after receiving training in early spring. The raters are professors in the education department of the college (one with the rank of administrative faculty and the other with the rank of tenure-track associate professor), and each has supervised student teachers in the teacher preparation program for several years. Both raters are

known to the participants because they work with them as instructors. The first rater has an earned Ph. D. in teacher education and has over 30 years of experience in teaching and educational leadership. The second rater is involved with recruiting and managing administrative tasks of record keeping for the degree-completion program, as well as adjunct teaching and supervising student teachers.

While it is ideal that both the researcher and the raters are unknown to participants, raters in this study were selected for their familiarity with the portfolio process (Gay, Arasian, & Mills, 2006). All 30 portfolios were rated using the *REPORT*, and scores for reflective writing were calculated for each of six domains in all 30 portfolios. Finally, reliability scores were calculated using the Pearson *r* correlation to determine interrater agreement (Gay, Mills, & Airasian, 2006).

In order to answer research question 2, the *REPORT* was utilized to compare two groups of portfolios, a control group (n=15) randomly selected from a cohort of portfolios created before the implementation of the instructional intervention (*Portfolio Reflective Writing Guide*) and a treatment group (n=15) randomly selected from portfolios created after instruction using the intervention. Differences between groups on each dimension and overall were calculated using *t* tests comparing the total scores of the two groups and a series of *t* tests for comparing the groups on each domain (Hinkle, Weirisma, & Jurs, 2003).

Finally, the answer to question 3 was determined in three ways. First, a criterion for the designation of high level of reflection was determined a priori as follows. On the *REPORT*, each domain may receive a total score of 9. Criteria for low, medium, and high levels of reflective writing were set so that scores of 0-3 per domain were designated low

level reflection, scores of 4-6 were designated medium levels of reflection, and scores of 7-9 were considered high levels of reflection. Second, the total number and percentage of portfolios that met the high reflection level was calculated for each group. Based on the research literature, no predetermined criterion was set for the expected percentage of each group to reach the highest level of reflection because any such criterion would be arbitrary (Rodgers, 2002). Finally, an independent samples Mann-Whitney *U* test was conducted to evaluate the hypothesis that the distributions of experimental and control groups as to levels of reflective thinking (low, medium, and high) would differ between groups across all six domains and for the total reflective level scores (Green & Salkind, 2008).

While no commonly-recognized criteria for high reflection exists in the literature (Rodgers, 2002), several studies reported results that guided the criteria selection for reflective writing in this study. In their examination of both process and product-oriented portfolios, Orland-Barak (2005) reported that no portfolios met the criteria for critical reflection (their highest level). They further reported that only 20% of the portfolios demonstrated reflection at the second level (dialogical) with an average of 6.6% per portfolio (Orland-Barak, 2005). Nagle (2009) reported more optimistic results in that four out of nine preservice teacher portfolios displayed reflective practice at the critical level. In both studies, all portfolios provided evidence of more low-level descriptive reflection than high-level reflection (Orland-Barak, 2005; Nagle, 2009). These results support findings by Delandshere and Arens (2003), that portfolios often present predominantly shallow descriptions of preservice teachers' reflective thinking.

Given the above analysis, for this study, any portfolio earning a high score (7-9) on at least two domains out of the six by both raters was considered sufficiently in-depth to contribute to preservice growth. While this may seem somewhat arbitrary, the strong consensus from the literature reveals that no single definition or measurement device exists to enable research on reflectivity (Rodgers, 2002; Robinson & Kelley, 2007; Tillema & Smith, 2007). In fact, it seems reasonable to assume that if portfolios display two out of six (33.3%) instances of high-level reflection as measured by two raters, then this demonstrates the candidates' ability to reflect deeply on their practice. In addition, it seems likely that across groups of preservice teachers in a program such as the one in this study, where optimal conditions for portfolio creation exist, at least half of their portfolios would be expected to meet this criterion for a high level of reflection.

### **Study Assumptions and Delimitations**

Several assumptions undergird this study. First, the study assumes that the portfolios involved are, in fact, created by the preservice teachers who are enrolled in the teacher education program and not the product of someone outside of the program. Second, the quasi-experimental cohort design assumes that the two cohorts are composed of subjects with similar interests, qualifications, and characteristics, though no pre-test comparison was conducted. The groups were compared using overall institutional GPA, and no significant differences were found. All participants must meet identical entrance requirements (2.5 GPA, basic skills tests, interviews, letters of recommendation, etc.) and share the goal of obtaining initial teacher certification in the area of early childhood. Finally, the study assumes that the dimensions of variation related to portfolio construction described in the setting section constitute favorable conditions for the

growth of reflective thinking, so any differences found in the levels of reflective thinking are assumed to be due to the study's treatment. History and maturity were not perceived as a threat as participants came from different cohorts.

In addition to study assumptions, several delimitations serve as boundaries for the research in this study. First, myriad educational institutions have adopted portfolio assessment, and portfolios exist in many content and subject areas. This study limits its investigation to the electronic, standards-based, exit portfolios of preservice teachers in a teacher education program at one small, liberal arts college. Second, in the field of education, portfolios have been utilized as an assessment tool in the early grades through the graduate level. In this study, the portfolios investigated are limited to elementary education majors at the undergraduate level. Finally, portfolios have been created for single courses and specific projects, as well as for purposes that encompass a wide range of competencies. The portfolio investigated in this study incorporates 12 beginning teacher standards organized into six domains. Therefore, it represents only a single kind of preservice teacher portfolio, the exit portfolio constructed for completion of the teacher education program. In view of these delimitations, study results may fail to reveal insights that would have come to light with a study of the far-ranging sample of portfolios found in the research literature.

### **Summary of Study Methodology**

This study utilized a quasi-experimental cohort design (King & Roblyer, 1984) to compare the levels of reflective thinking in portfolios of preservice teachers who have and have not received an intervention to teach portfolio reflective writing. Two groups (n=15 each) of preservice teachers were randomly selected from a population of



preservice teachers. Portfolios from these participants were assessed using a rubric to measure reflective thinking and the results were compared using  $t$  tests. Further, the two groups were compared with respect to criterion set for the numbers (percentages) of preservice teachers reaching high levels of reflection, and the percentage of portfolios receiving the designation of high-level reflection was calculated. Finally, an independent samples Mann-Whitney  $U$  test was conducted to evaluate the hypothesis that the distribution of levels of reflective thinking (low, medium, and high) would differ across all six domains and for the total reflective level (Green & Salkind, 2008). These methods and procedures provide data to answer the three research questions posed in this study.

## Chapter 4: Results

### Overview of the Study

This study was designed to determine whether teacher portfolios can be validly and reliably assessed, to investigate the effect of an instructional tool on increasing the level of reflective thinking in elementary preservice teachers' portfolios, and to find whether electronic portfolios designed and assessed in optimal conditions represent sufficient quality to make them useful in practice. A quasi-experimental cohort design (King & Roblyer, 1984) was used to compare the reflective thinking contained in the developmental portfolios of two groups of elementary preservice teachers. Two groups (n=15 each) of preservice teachers were randomly selected from a population of preservice teachers. After receiving training, two raters assessed portfolios from these participants using a rubric to measure reflective thinking (*REPORT*). The results were compared using a series of *t* tests. Further, the two groups were compared with respect to a criterion set for the numbers (percentages) of preservice teachers reaching high levels of reflection. To determine this, the percentage of portfolios receiving the designation of high-level reflection was calculated. Finally, an independent samples Mann-Whitney *U* test was conducted to evaluate the hypothesis that the distribution of levels of reflective thinking (low, medium, and high) would differ across all six domains and for the total reflective level (Green & Salkind, 2008).

As described in Chapter 1, this study sought to answer the following research questions:

1. Does the *Rubric for Evaluating Portfolio Reflective Thinking (REPORT)* demonstrate sufficient validity and reliability for use in measuring reflective thinking in preservice teacher portfolios?
2. Do levels of reflective thinking in preservice teacher portfolios, as measured by the *REPORT*, differ between students who have and have not received instruction using a Scaffolding Intervention Tool?
3. Do elementary preservice teachers' portfolio rationale statements and reflective essays, as measured by the *REPORT*, show sufficient depth of reflective thinking to aid their growth as teachers?

Study findings are reported in this chapter. Results are given both in terms of preliminary findings on the population and in terms of findings on each research question.

### **Preliminary Findings on Sample and Instrumentation**

**Sample characteristics.** The sample used in this study was randomly selected from the population of undergraduate students with a major in elementary education who completed their degrees between 2007 and 2010. The population from which the sample was drawn consisted mainly of female elementary education teacher candidates. This study sample includes only one male, but since only 3% of the total population of elementary education majors at this institution is male, this number was aligned with population characteristics.

Participants for the control group (n=15) were randomly drawn from the population of elementary education teacher candidates who completed their degree between May 2007 and December 2009 and who had not received any specific training

on how to write reflective statements. Participants for the treatment group (n=15) were randomly drawn from the population of elementary education teacher candidates who completed their degree in the spring of 2010 and who had received the study intervention. This treatment provided training in how to write reflective statements; including questions and prompts to assist reflective thinking and exposed to a detailed rubric with examples of reflective statements (see Appendix A).

Because this study involved a post-test-only design, the overall institutional grade point average (GPA) for both groups was used to compare groups for ability levels. An independent-samples *t* test was conducted to evaluate the hypothesis that there is no difference in GPA between the control group and the treatment group in this study. The test was not significant ( $t(28)=-.40, p=.69$ ) thereby confirming the hypothesis that there was no significant difference in ability levels between the two groups. GPA between candidates in the control group ( $M=3.58, SD=.29$ ) and the treatment group ( $M=3.73, SD=1.45$ ) was not significantly different, indicating that any differences in ability (as indicated by GPA) between the two groups was due to chance.

**Instrument characteristics.** Content validity for the researcher-created rubric (*REPORT*) was demonstrated through expert analysis and verification. An early draft of the *REPORT* was sent in the fall of 2009 to eight experts recognized for their expertise in portfolio assessment in teacher education through research published in peer-reviewed articles. Each expert was asked to evaluate the content of the rubric, as well as the descriptions of levels of performance, sample reflective statements, and scoring guide. Of the eight experts, two responded with comments indicating that the rubric was grounded in the research and the descriptors appeared to measure the construct of reflection well.

Further, a draft of the *REPORT* was provided to the raters and other colleagues in the teacher preparation program who are familiar with portfolio procedures and feedback was used to clarify the descriptors. Raters indicated that both the descriptors for both types of reflection and levels of quality were clear and easy to use.

In order to complete preliminary interrater reliability calculations, two raters were trained on how to use and score the *REPORT*. Both raters were full-time professors in the teacher education department that served as the setting for this study. One rater has earned a Ph.D. in teacher education and the other an M.Ed. in educational leadership. Both raters regularly serve as college supervisors for student teachers in the early childhood programs in the institution where the study took place. Participants are known to both raters, which may increase rater bias and reduce interrater reliability, as noted in the limitations.

In addition, both raters attended the portfolio intervention sessions where the teacher candidates were trained in using the rubrics and the prompts. Additional scoring directions were provided orally for the raters, and any questions they had were answered by the researcher during the scoring process. Initial training revealed some confusion about whether the scoring should be approached holistically or analytically. In holistic scoring, a single score is assigned to one product based on several characteristics. Analytic scoring refers to assigning a separate score to each specific criterion. Raters were instructed to score the *REPORT* holistically within each domain, as recommended for portfolio assessment in the research literature (Johnson, Mims-Cox, & Doyle-Nichols, 2006; Meeus, Petegem, & Engels, 2009). Discussion during interrater training clarified

the confusion, and the two raters maintained a similar holistic approach to scoring within each domain throughout the scoring process.

After the trainings were completed, both raters scored 10 portfolios drawn from the archives. Interrater reliability was computed using a Pearson  $r$  correlation (Gay, Mills, & Airasian, 2006), with the goal for interrater agreement to reach  $r=.80$  or higher. This represented an optimistic goal for a performance rubric such as a portfolio. While standardized achievement tests should have high reliability (greater than .90), for measures of psychological constructs such reflective thinking, more moderate reliability scores for other types of measures may satisfy the researcher (Gay, Mills, & Airasian, 2006).

Initial Pearson  $r$  correlations were computed for the first 10 practice portfolios between raters 1 and 2, and the result was  $r=.72$ , which was significant at the  $p \leq .05$  level (2-tailed). This represents moderate interrater reliability but did not reach the preset criteria of  $r = .80$  that was desired. The researcher conducted additional training and the raters discussed discrepancies in scoring for each domain. Raters then scored two additional portfolios drawn from the archives. Results from the additional two portfolio scores were much closer, though an additional Pearson  $r$  correlation was not conducted due to the small sample size of two. Scores were judged sufficiently close together for the researcher to allow the raters to move on to score the study sample.

### **Findings on Research Question 1**

Research Question 1 asked if the *Rubric for Evaluating Portfolio Reflective Thinking (REPORT)* demonstrated sufficient validity and reliability in practice for use in measuring reflective thinking in preservice teacher portfolios. Since the domains and

components were derived from research on teacher portfolios and were confirmed by expert review by recognized practitioners in the field, results indicated that the rubric has sufficient validity in terms of content.

The *Rubric for Evaluating Portfolio Reflective Thinking (REPORT)* used in this study was created by the researcher so no published interrater reliability data were previously available. The *REPORT* was organized into six sections, corresponding with the six domains contained in the preservice teacher portfolios. Reflective writing scores were calculated for each domain by summing the scores on each of three types of reflection (technical/descriptive, personal growth, dialogic/critical) and a total score was calculated. Pearson  $r$  correlations were computed for each domain and for the total score and are listed in Table 1.

Table 1

*Pearson  $r$  Interrater Reliabilities for REPORT Ratings across Domains*

Domain Names	$r$
Domain A: Knowledge	.51
Domain B: Planning	.63
Domain C: Instruction	.50
Domain D: Assessment	.61
Domain E: Classroom Environment	.62
Domain F: Professional Growth	.55
Total	.66

The total Pearson  $r$  (.66) was moderate and did not reach the desired level of .80. Pearson  $r$  scores for the individual domains ranged from .50 (Domain C: Instruction) to .63 (Domain B: Planning), a narrow range. Reliability coefficients reported for the individual domains were generally lower than for the total. According to Gay, Mills, and

Airasian (2006) it is reasonable for researchers to expect lower reliability coefficients on individual subtests than on total test scores due to the smaller number of items. Scores on the individual domains serve as subtests on the instrument used in this study, the *Rubric for Portfolio Reflective Thinking (REPORT)*.

## **Findings on Research Question 2**

Research Question 2 asked if levels of reflective thinking in preservice teacher portfolios, as measured by the *Rubric for Evaluating Portfolio Reflective Thinking (REPORT)*, differed between students who had and had not received instruction using an instructional intervention tool called the *Portfolio Reflective Writing Guide*. Aimed at increasing preservice teachers' in-depth reflective analysis, the guide contained the *REPORT* rubric, model exemplars of reflective statements, and guiding questions to prompt deep thinking. Data to address this question were *REPORT* rubric ratings from the two raters on 30 teacher portfolios: 15 from the experimental group, who received the instruction in reflective thinking, and 15 in the control group, randomly selected from a cohort of students who graduated between May 2007 and December 2009.

Ratings of the two raters were averaged and *t* tests were computed using the total scores and the scores on each domain. A series of independent-samples *t* tests was conducted to evaluate the hypothesis that teacher candidates who have received an instructional intervention in reflective thinking write more in-depth reflective statements than candidates who have not received this instructional intervention. Means and standard deviations for each domain and the total scores are presented in Table 2.



Table 2

*Results of t-test Comparisons of Experimental and Comparison Groups*

Domain	Group (n=15)	Mean	SD	df	<i>t</i>	Sig. (2-tailed)
A	Treatment	6.4	1.59	28	2.27	.03*
	Control	5.3	1.00	28	2.27	
B	Treatment	5.9	1.58	28	1.35	.19
	Control	5.2	1.24	28	1.35	
C	Treatment	5.7	1.41	28	2.58	.02*
	Control	4.4	1.35	28	2.58	
D	Treatment	5.8	1.64	28	2.74	.01**
	Control	4.4	1.11	28	2.74	
E	Treatment	6.1	1.51	28	3.59	.00**
	Control	4.2	1.33	28	3.59	
F	Treatment	5.9	1.52	28	3.08	.01**
	Control	4.5	1.04	28	3.08	
Total	Treatment	35.8	8.55	28	2.83	.01**
	Control	28.0	6.33	28	2.83	

---

\* $p < .05$ . \*\* $p < .01$

Results indicate that the treatment group, whose members had received instruction, scored significantly higher than the control group, whose members had not received instruction, using an alpha level of  $p < .01$  on the total *REPORT* score on three of the six domains (Domain D: Assessment, Domain E: Classroom Environment, and Domain F: Professional Growth). Using the alpha level of  $p < .05$ , the treatment group scored significantly higher than the control group on two of the six domains (Domain A: Knowledge and Domain C: Instruction). Scores for the treatment group were not significantly different at either level for only one domain (Domain B: Planning).

### **Findings on Research Question 3**

Research Question 3 asked if elementary preservice teachers' portfolio rationale statements and reflective essays, as measured by the *Rubric for Evaluating Portfolio Reflective Thinking (REPORT)*, showed sufficient depth of reflective thinking to aid their growth as teachers. The assumption underlying this question is that when preservice teachers understand the purpose of the portfolios, have been provided with sufficient training and support to reflect deeply on their practice, and have been provided with assessment criteria and examples, that they will be able to write reflective statements and essays that demonstrate high levels of reflection. For this study, scores for each portfolio in both groups from both raters on each domain were averaged and assigned to one of three levels reflection: low (0-3), medium (4-6), and high (7-9). Portfolios that met preset criteria of a high level of reflection on at least two domains out of six were designated as posting sufficiently high levels of reflection to aid their growth as teachers. Though no a priori criterion was set for the number and percentage of each group to reach the highest level of reflection, it was expected that a minimum of 50% of the portfolios in the treatment group would meet the highest level of reflection, since they had experienced optimal conditions for creating portfolios (Zeichner & Wray, 2001). Results indicate that 47% of the portfolios in the treatment group met the criteria for sufficiently high levels of reflection while only 6.7% met these criteria in the control group. Table 3 below presents the numbers of portfolios scoring at the high level of reflection.

Table 3

*Number and Percentage of Portfolios Reaching High Levels of Reflection*

# out of 6 Group <sup>a</sup> :	1	2	3	4	5	6	Total $\geq 2/6$
Treatment	1 (7%)	2 (13%)	1 (7%)	1 (7%)	-	3 (20%)	7 (47%)
Control	1 (7%)	1 (7%)	-	-	-	-	1 (7%)

<sup>a</sup>n=15

These numbers show that the treatment group contained considerably more portfolios meeting the preset criteria of two out of six domains reaching the highest level of reflection. While the percentage of the treatment group that did meet the criteria for high levels of reflection (47%) was just short of the expected 50%, the treatment appears to have increased the percentage of candidates who are capable of high levels of reflection. It is noteworthy that three candidates in the treatment group (20%) earned scores reaching the highest level of reflection in all six domains (100%).

To further clarify the answer to research question 3, an independent-samples Mann-Whitney *U* test was conducted to evaluate the hypothesis that the distribution of levels of reflective thinking (low, medium, and high) would differ across all six domains and for the total reflective level (Green & Salkind, 2008). The results of the test were in the expected direction and significant for three out of six domains (knowledge and professional growth) and for the total reflective level. Table 4 presents the results of the Mann-Whitney *U* test comparison.

Table 4

*Mann-Whitney U Comparison of Reflection Levels by Domain*

Domain Names	Sig.	Decision
Domain A: Knowledge	.01**	Reject Null
Domain B: Planning	.09	Retain Null
Domain C: Instruction	.04*	Reject Null
Domain D: Assessment	.06	Retain Null
Domain E: Classroom Environment	.21	Retain Null
Domain F: Professional Growth	.02*	Reject Null
Total	.00**	Reject Null

\* $p < .05$ . \*\* $p < .01$

The results of the Mann-Whitney  $U$  further demonstrate that preservice teachers with instructional intervention in reflective writing can demonstrate their own development in the areas of knowledge, instruction, and professional growth using more in-depth analysis than preservice teachers who have not had this instruction. Further, preservice teachers who had received the instructional intervention reached higher levels of reflection overall. In addition, the Mann-Whitney  $U$  results triangulate the findings from the  $t$  tests used to answer Research Question 2.

### Summary of Study Findings

In summary, an analysis of the researcher-created portfolio assessment instrument used in this study (*Rubric for Evaluating Portfolio Reflective Thinking (REPORT)*) revealed moderate levels of interrater reliability between the two portfolio raters after training. Feedback from expert reviewers and raters indicated that the *REPORT* demonstrated sufficient content validity to be used to measure reflective thinking in preservice teacher portfolios.

Next, on five of the six domains and on the total score, the treatment group, which had received instruction in reflective writing, scored significantly higher than the control

group, which had not received instruction. There was no significant difference between groups on the Planning domain.

Finally, an analysis of the overall levels of reflection in the portfolios of both groups showed that a substantially higher percentage of preservice teachers in the treatment group (47%) wrote reflective statements that reached high levels of reflection than the percentage of preservice teachers in the control group (6.7%). Mann-Whitney *U* comparison results indicate that the group that received specific instruction in how to write reflectively demonstrated significantly higher levels of reflection on three out of the six domains and on the total reflective level score. The interpretations of, and implications for, these findings will be explored in the final chapter.

## Chapter 5: Discussion

### Overview of Study Discussion

Criticisms of teachers and teacher education have reached a crescendo resulting in a cacophony of cries for highly-effective teachers and improved teacher training programs (Darling-Hammond & Youngs, 2002). The demands of policymakers, taxpayers, and educational leaders for increased accountability have galvanized the educational community around a push for excellence (Stronge, Ward, Tucker, & Hindman, 2007). Stronge et al. (2007) succinctly summarized this current state of affairs when they stated, “Most recently, reauthorization of the Elementary and Secondary Act, better known as the *No Child Left Behind Act*, is intended to tie federal education funding directly to improvements in student test scores” (p. 166).

Consequently, colleges of teacher education and various alternative teacher preparation programs have designed a variety of performance assessments to ensure that their graduates can meet rigorous performance standards. Increasing pressure on states to include student achievement scores in their teacher evaluation procedures ripples out to teacher preparation programs as they present newly-minted teachers to the profession.

Though few studies tie preservice teacher portfolios directly to high levels of K-12 student achievement, various types of portfolios have become ubiquitous in colleges of teacher education at both the undergraduate and graduate levels. In the wake of the push for teacher education reform, performance assessments have been used in addition to standardized tests to measure preservice teacher quality (Darling-Hammond & Youngs, 2002). Preservice teacher portfolios are one vehicle frequently employed to develop and document preservice teachers’ reflective capabilities. The claim is often

made that portfolios demonstrate teacher candidates' growth over time and encourage reflective thinking, thus ensuring high quality teaching (Milman, 2005; Riedinger, 2006; Ring & Foti, 2006).

This study was designed to determine whether teacher portfolios can be validly and reliably assessed, to investigate the effect of an instructional tool on increasing the level of reflective thinking in elementary preservice teachers' portfolios, and to find whether electronic portfolios designed and assessed in optimal conditions represent sufficient quality to make them useful in practice. Presumably, teachers who can reflect deeply on their work and its impact on others can improve the quality of their teaching. In order to determine the impact of an instructional intervention, it is necessary to assess the level of reflective thinking in portfolios before and after such an intervention. Because no instrument currently exists to measure the depth of reflection in preservice teacher's electronic portfolios, this study offered a rubric to measure portfolio reflective thinking and used that rubric to measure reflective thinking after instruction had taken place.

This study contributed to the extant literature on portfolio evaluation by testing an assessment instrument to measure reflective thinking in portfolios and by examining the effects of a scaffolding intervention on the levels of reflection in undergraduate, elementary preservice teachers' electronic, standards-based, exit portfolios. This study focused on finding the answers to the following research questions:

1. Can a research-based instrument be designed that teacher educators can use to measure reflective thinking in practice?

2. Can an instructional intervention designed to scaffold reflective thinking increase the levels of elementary preservice teachers' reflective thinking in the electronic portfolio rationale statements and reflective essays?
3. Do elementary preservice teachers' portfolio rationale statements and reflective essays show sufficient depth of reflective thinking to aid their growth as teachers?

**Overview of study methodology.** This study utilized a quasi-experimental cohort design (King & Roblyer, 1984) to compare the levels of reflective thinking in portfolios of preservice teachers who had and had not received an intervention to teach portfolio reflective writing. Two groups (n=15 each) of preservice teachers were randomly selected from a population of preservice teachers. Portfolios from these participants were assessed using a rubric to measure reflective thinking, and the results were compared using *t* tests. Further, the two groups were compared with respect to a criterion set for the numbers of preservice teachers reaching high levels of reflection as indicated by the numbers of portfolios receiving the designation of high-level reflection. Finally, an independent samples Mann-Whitney *U* test was conducted to evaluate the hypothesis that the distribution of levels of reflective thinking (low, medium, and high) would differ across all six domains and for the total reflective level (Green & Salkind, 2008). This chapter discusses the implications of the findings for each research question, offers implications for practice, and explores suggestions for future research in preservice teacher portfolio assessment.



## Interpretation of Findings

**Research question 1 findings and interpretations.** Research Question 1 investigated whether the *Rubric for Evaluating Portfolio Reflective Thinking (REPORT)* demonstrated sufficient validity and reliability in practice for use in measuring reflective thinking in preservice teacher portfolios. Since the domains and components were derived from research on teacher portfolios and were confirmed by expert review of recognized practitioners in the field, results indicated that the rubric has sufficient validity in terms of content to be useful to teacher education personnel as they score preservice teachers' standards-based, electronic, exit portfolios.

Interrater reliability was determined by calculating Pearson  $r$  scores for both the individual domains and the total portfolio score on the *REPORT*. Scores on the individual domains ranged from .50 (Domain C: Instruction) to .63 (Domain B: Planning), which is a narrow range. The total Pearson  $r$  (.66) was moderate and did not reach the desired level of .80. This moderate level of interrater reliability indicates that, even with training, rater agreement is difficult to achieve using a scoring rubric to assess portfolios. Gay, Mills, and Airasian (2006) indicated that performance assessments often post scorer agreement rates lower than those of standardized tests; the latter can reach levels as high as  $r=.90$ .

However, the Pearson  $r$  of .66 reported in this study is within the range of interrater agreement levels reported in several other studies. On the low end, a study conducted by Derham and Diperna (2007) posted scores of interrater reliability with “a median kappa coefficient of 0.14 across all raters” (p. 373). In contrast, Sulzen and Young (2007) achieved relatively high levels of interrater reliability (0.85) in their study

of portfolios created early in a preservice teachers' program. In addition, raters in this study reported the average time required to score each portfolio was 15 to 20 minutes, which they felt to be a reasonable period of time to dedicate to an assessment. This relatively short time increased the likelihood that reliable assessment would realistically occur. However, portfolios created early in a teacher education program would not be useful for comprehensive assessment of standards-based teacher capability desired at program exit.

Johnson (2006) reported high levels of reliability and construct validity when scoring online portfolios in an engineering education program. However, she also reported that interrater reliability as measured by Pearson  $r$  correlations improved over 3 years, and adjudicated correlations for each scoring dimension ranged from 0.80 to 0.88 by the third year. (Johnson reported that when scores differed by more than two points on any item in their portfolio rubric, scores were adjudicated by discussion between raters until agreement was reached.) However, correlation coefficients for the first year were much lower, ranging from 0.28 to 0.64. Clearly, in Johnson's study, further experience with the scoring rubric and process increased interrater reliability over time. The study also reported high coefficients of determination and determined that the construct the portfolio was measuring (technical communication) "accounted for no less than 73% of the variance within the model" (Johnson, 2006, p. 283).

With respect to interrater reliability, this study supports the conclusion that, even with rater training, high levels of interrater agreement are difficult to achieve in portfolios. Multiple trainings may be necessary over several years of a portfolio's development to produce reliability levels sufficient to ensure valid interpretations of

teacher reflection. Further, since raters reported satisfaction with the clarity of the *REPORT*, it seems likely that continued use of such an instrument over several years would enhance interrater agreement. The addition of a detailed written scoring guide would aid raters and ensure consistent scoring approaches across portfolios. Measures such as these would increase the interrater reliability. Also, following the example of the Johnson (2006) study, raters could adjudicate scores through discussing any discrepancies until agreement is reached.

**Research question 2 findings and interpretations.** Research Question 2 explored whether levels of reflective thinking in preservice teacher portfolios, as measured by the *Rubric for Evaluating Portfolio Reflective Thinking (REPORT)*, differed between students who had and had not received instruction using an instructional intervention tool called the *Portfolio Reflective Writing Guide*. Aimed at increasing preservice teachers' in-depth reflective analysis, the guide contained the *REPORT* rubric, model exemplars of reflective statements, and guiding questions to prompt deep thinking. During this training, preservice teachers created a sample reflective statement and worked with a partner to evaluate the quality of this statement using the *REPORT*. Peer discussion and support was intended to enable preservice teachers to go beyond the shallow reflections often reported in the research literature (Delandshere & Arens, 2003; Orland-Barak, 2005).

Results indicate that the treatment group, which had received instruction, scored significantly higher than the control group, which had not received instruction (alpha level of  $p < .01$ ) on the total *REPORT* score and on three of the six domains (Domain D: Assessment, Domain E: Classroom Environment, and Domain F: Professional Growth).

At the alpha level of  $p < .05$ , the treatment group scored significantly higher than the control group on two of the six domains (Domain A: Knowledge and Domain C: Instruction). Scores between the treatment and control groups were not significantly different at either level for only one domain (Domain B: Planning).

While the treatment group posted higher reflective writing scores overall, these data bear closer examination by domain. Results support the conclusion that preservice teachers are better able to reflect deeply on some types of performance than others. For example, three domains (assessment, classroom environment, and professional growth) showed significant differences at the  $p < .01$  alpha level, while two domains (knowledge and instruction) reported significant differences at the  $p < .05$  alpha level. The planning domain showed no significant difference.

The study research methodology and results do not provide a clear answer to why this might be the case, but evidence from the research literature may illuminate interpretation of these findings. Sulzen (2007) concluded that faculty members made strong judgments in the areas of “content knowledge, general pedagogic knowledge, instructional design, instructional delivery, assessment, and reflection” (p. 115). These findings seem consistent with the finding in the current study that preservice teachers in the treatment group posted significantly higher reflective scores in several similar areas: knowledge, instruction, assessment, and professional growth. While Sulzen’s (2007) investigation focused on faculty raters’ judgments and this study looks at preservice teachers’ reflective statements, there may be something about the nature of content knowledge, instruction, assessment, and professional growth that lends itself to high levels of reflection.

The finding of no significant difference in the area of planning is interesting and poses a challenge to interpretation. One possible explanation for this finding is that the teacher education program under investigation in this study has developed a strong focus on planning for many years. Portfolios created by members of the control group would have received detailed coursework and training in both lesson-planning and unit-planning procedures. The instructional intervention in this study did not highlight planning, specifically, so this may be one area of teacher performance in which the control group and the treatment group had experienced similar levels of preparation, despite the use of the *Portfolio Reflective Writing Guide*. Preservice teachers in both groups may have been more comfortable with writing reflectively about planning, since they had multiple opportunities to write lesson plans and units. Most of the lesson and unit planning assignments contained in the courses require a section for reflection after the lesson or unit has been taught. This has been true for quite a few years in the teacher education program under study.

In summary, the treatment group who had undergone specific instruction in reflective thinking did benefit significantly from portfolio-specific instruction in how to demonstrate reflection in clear and convincing ways, at least in some domains. This study is able to conclude that training and instruction in writing reflectively appears to be important to helping elementary preservice teachers demonstrate their reflective capability in standards-based exit portfolios.

**Research question 3 findings and interpretations.** Research Question 3 asked if elementary preservice teachers' portfolio rationale statements and reflective essays, as measured by the *Rubric for Evaluating Portfolio Reflective Thinking (REPORT)*, showed

sufficient depth of reflective thinking to aid their growth as teachers. Results indicate that 47% of the portfolios in the treatment group met the criteria for sufficiently high levels of reflection while only 6.7% met these criteria in the control group. These numbers show that the treatment group contained considerably more portfolios meeting the preset criteria of two out of six domains reaching the highest level of reflection. While the percentage of the treatment group that did meet the criteria for high levels of reflection (47%) was just short of the expected 50%, the treatment appears to have increased the percentage of candidates who are capable of high levels of reflection. It is interesting to note that three candidates in the treatment group (20%) earned scores reaching the highest level of reflection in all six domains (100%).

To further clarify the answer to research question 3, an independent samples Mann-Whitney  $U$  test was conducted to evaluate the hypothesis that the distribution of levels of reflective thinking (low, medium, and high) would differ across all six domains and for the total reflective level (Green & Salkind, 2008). The results of the test were in the expected direction and significant for three out of six domains and for the total reflective level.

The results of the Mann-Whitney  $U$  further demonstrate that preservice teachers with instructional intervention in reflective writing can demonstrate their own development in the areas of knowledge, instruction, and professional growth using more in-depth analysis than preservice teachers who have not had this instruction. Further, preservice teachers who had received the instructional intervention reached higher levels of reflection overall. In addition, the Mann-Whitney  $U$  results triangulate the findings from the  $t$  tests used to answer research question 2.

This study did not specifically investigate reasons why reflection may differ by domain in the standards-based, exit portfolios. However, it is reasonable to offer hypotheses regarding these findings, based on the structure of the portfolio itself. Experience with scoring portfolios, as part of the regular exit assessment protocol for the college where this study took place, suggested that the knowledge domain contained a greater number of artifacts per student than the other five domains. Since there were more reflective statements available to serve as evidence of high levels of reflection, it seems likely that raters were able to ascertain more in-depth reflection for this area.

In contrast, the domain of professional growth usually contains the fewest number of artifacts. In fact, teacher candidates often struggle to obtain artifacts for this section and rely on descriptions in their reflective statements to augment the actual attached documents. Insufficient data are available at this time for the researcher to hypothesize about the reasons for significant differences between groups on the professional growth section of the portfolio.

Whatever the reasons, nearly half of the preservice teachers performed at high levels of reflection, including critical reflection engaging multiple perspectives, social justice issues, and moral concerns. Further, since analysis of the Mann-Whitney  $U$  results indicated that the general distribution of reflection scores *across reflection levels* was significantly higher for the treatment group on three domains and the total portfolio score, it is reasonable to conclude that training and support can increase levels of reflection, even if a large percentage of portfolios do not reach the very highest level of reflection. As with any measure of performance, variation across portfolio reflection is expected. However, if teacher education programs embed instruction regarding reflective writing

throughout their programs, it is likely that over time most preservice teachers will be capable of reflecting deeply on their work, demonstrating that reflection in their portfolios and enhancing their growth as effective practitioners. Considering that the review of the research literature unearthed many studies posting low critical reflection (Delanshere & Arens, 2003; Orland-Barak, 2005), findings from this study constitute evidence that high levels of reflection are obtainable, with sufficient training and support.

### **Implications for Practice**

**Teacher education personnel can create valid and reliable rubrics for portfolios.** Though findings from this study indicate that interrater reliability is a challenge to achieve, it is possible to design a clear rubric that measures the construct of reflection validly and can be used reliably by teacher education practitioners. The *REPORT* in this study was used for research purposes only and was not the actual rubric used in the teacher education program to evaluate preservice teacher portfolios. This study implies that it would be beneficial to revise the *REPORT* and use it to score portfolios in the teacher education program described in this context. Both raters indicated that the reflection categories and levels of quality were clear, the sample reflective statements were helpful, and criteria for numbers of artifacts and quality of writing conventions could easily be added to make this the actual rubric used (Yao, Aldrich, & Foster, 2008).

Once a teacher preparation program develops a rubric, it will be essential that it provide extensive training to ensure that the raters understand the constructs and scoring procedures, as well as to increase the likelihood of high interrater reliability. Utilizing many raters would strengthen validity and reliability and help to ensure valid decisions



about preservice teacher reflective capability. Training should include the opportunity for raters to compare notes (adjudication), time to engage in detailed discussion regarding any discrepancies in ratings, and analysis of any interrater reliability calculations. State and national accreditation guidelines call for teacher education programs to demonstrate evidence that their unit assessment systems are fair and without bias. Interrater training such as that described here would meet this requirement. Since colleges of teacher education often use adjuncts to supervise student teachers, both adjuncts and full-time professors would benefit from rigorous training procedures in this area. Further, interrater reliability would probably increase over time as raters gained practice using the scoring rubric (Johnson, 2006).

**In-depth reflection can be developed using training and prompts.** Since the treatment group posted significantly higher reflection scores on five out of six domains and on the total score, study results indicate that it is possible for other teacher education programs to help preservice teachers produce reflective writing using instruction and prompts. Specifically, training and support, including a clear rubric and examples, enable preservice teachers to create reflective rationales and essays that provide full explanations of their work. The *REPORT* used in this study delineated three types of reflection with levels of quality for each one that seemed to guide preservice teachers as they constructed their portfolios. Teacher education program design and coursework that includes specific scaffolding for reflective thinking and writing is more likely to enable creation of rich portfolios that contain greater levels of critical reflection than teacher education curriculum that omits such training.

Though this study measured preservice teacher portfolios at only one checkpoint (program exit), it would be reasonable to hypothesize that instruction and training built into courses throughout the teacher education program would further enhance reflective capabilities in preservice teachers. In fact, survey results from a pilot study that was not included in this report indicate that teacher candidates desired to incorporate portfolio artifacts and reflection into their coursework throughout their program (Pennington, 2009). Further, results suggest that the *REPORT* developed for this study could serve as a useful formative assessment tool. Teacher candidates may benefit from using it to evaluate portfolio drafts, either alone or in discussions with peers, which is a practice supported in the research literature (Gordinier, Conway, & Journet, 2006). Discussions with peers and professors also provide teacher candidates with the opportunity to demonstrate reflective capability orally, a skill that will serve them well during employment interviews. Preservice teachers persist in their perception that portfolios are, and should be, beneficial for employment, though that is rarely the case (Barrett & Wilkerson, 2004; Ntuli, Keengwe, & Kyei-Blankson, 2009; Zeichner & Wray, 2001). Even if principals do not desire to examine portfolios during the interview process, the practice of reflection found in portfolio construction may serve to strengthen critical thinking skills as candidates experience the hiring process (Amobi, 2006; Gordinier, Conway, & Journet, 2006).

**Highly reflective portfolios require time and effort.** The first two implications for practice described above are directly connected to the research questions and findings in this study. Underlying such implications are concerns of a more subtle nature. Ultimately, teacher education programs need to answer the question of value: Are

portfolios worth the investments of time and effort that are necessary as a foundation for sound assessment practice? Shulman's (1998) poignant comment still pertains,

Portfolios done seriously take a long time. They are hard to do. Teaching is a job that occupies every waking and some nonwaking moments of good teachers.

(Some of those nonwaking moments are at night, some while teaching.) Given such demands, the question is: Is that much work worth it? (p. 35).

Though that is a question each teacher education program must answer in light of its own values and available support, the implications that portfolios can be validly and reliably scored and that training can produce high levels of reflection, offer strong support for making the decision to invest the time and effort required.

### **Study Limitations**

Every research study has limitations, and this one is no exception (Patten, 2005). First, the single setting may limit generalizability to other teacher education institutions that are similar in size, mission, and population characteristics. Second, the small sample size (n=30) may reduce the confidence that would be available from using a sample size larger than 100. Third, selection threats due to subject characteristics may distort the differences between groups. Although the sample groups were randomly drawn from two populations assumed to be similar, and when groups were compared using overall institutional GPA, no significant differences were found, they are still considered intact groups (Patten, 2005). Though groups appeared to be similar in ability levels, one group may contain individuals with attributes related to reflectivity or other characteristics which may be the cause of any differences found, rather than the differences being caused by the treatment.

Fourth, the study is subject to threat from researcher bias. Both the researcher and raters are instructors in both early childhood programs and the participants are known to them. The portfolios contained in *LiveText* show the names of their creators and it is not possible to remove the names for a blind rating; therefore, prior knowledge and experience with the participants may influence portfolio ratings and conclusions drawn from the data.

Finally, a threat to validity from history or instructional factors other than the specific treatment may serve to offer one group some additional assistance with writing reflective statements. In fact, this study is subject to a selection-history interaction internal threat to validity (Patten, 2005). The sample of students who served as the treatment group (Spring 2010 student teachers) may have systematically experienced instruction in the program that was not part of the treatment, and it was this instruction that led them to different levels of reflection. For example, if the college supervisors for the treatment cohort all spent additional time giving detailed feedback to participants regarding the nature and level of reflection in their reflective statements, this could have accounted for any differences in the post measure (Patten, 2005). Readers wishing to generalize the results of this study would need to conduct further study in their own context.

### **Implications for Future Research**

There are four primary recommendations for further research that grow out of this study. Study results point to several areas not directly investigated in this study.

**Research on portfolio assessment is needed using larger contexts.** The first recommendation is a call for larger-scale studies to investigate increased methods of valid

and reliable portfolio assessments. Studies from large teacher education programs that train and utilize many raters and conduct interrater reliability calculations and studies that develop and hone sound instruments, would contribute to the knowledge base and serve teacher educators as they prepare the nation's future teachers. Specifically, rubrics such as the researcher-developed *REPORT* designed for use in this study would enable valid, reliable, and timely scoring of portfolios for a variety of purposes.

**Research is needed to link highly reflective portfolios to effective classroom performance.** A second recommendation for further research is that studies are needed to clarify the relationship between constructs such as teacher reflective capability displayed in portfolios and excellent teacher performance. Yao and Thomas, et al. (2008) called for research to validate the use of portfolios for reflection. This study aims to answer this call but further research is needed with larger sample sizes and in a variety of institutions and contexts. Further, research that establishes a direct link between portfolios and teacher quality would strengthen the claim that portfolios enhance excellent performance.

**Research is needed to link highly reflective portfolios to increased K-12 student achievement.** A third recommendation for further study is that portfolio assessment needs to be linked to K-12 student learning outcomes. Impact on K-12 student learning seems to be the gold standard called for by state and national accrediting bodies (Carney, 2006; Gathercoal, Love, & McKean, 2007). Even if portfolios can document high levels of reflective writing, the claim that in-depth reflection enhances teacher performance in ways that increase student achievement needs to be substantiated with outcome data (Zeichner & Wray, 2001).

A serious paucity of research on this connection between reflection and student performance currently exists in the body of literature. Very few studies have investigated how *preservice* teacher portfolios might serve as evidence of impact on K-12 student learning (Levitt & Schreiber, 2008). Using student teacher reflective statements, Levitt and Schreiber (2008) concluded that student teachers collected a variety of types of evidence for student learning across all subject areas. The authors reported that teacher candidates did not generally link the data they provided with claims and inferences made regarding student achievement. Further, they gathered no external data regarding student achievement. Given the intense focus on accountability and the need for teacher educators to demonstrate impact on student learning, empirical evidence from further research would provide empirical evidence that highly reflective portfolios allow teacher candidates to improve student achievement.

**Reflection may be better assessed using a hermeneutic paradigm.** A final recommendation for further research is to study alternate qualitative and hermeneutic methods of portfolio assessment not rooted in quantitative standards for reliability and validity. This recommendation acknowledges the inherent tension in portfolio evaluation between validity and reliability (Barrett & Wilkerson, 2004; Wray, 2008). The paradigm conflict in portfolios that pits summative documentation of high quality performance with formative documentation of growth and reflection is heightened when psychometric guidelines for measurement are applied to portfolio rubrics, as was done in this study.

It may be possible to strike a balance between the desire for sound assessment practices and the need for solid empirical research (Carney, 2007). Further development of instruments such as the *REPORT* used in this study may provide reliable instruments

that encourage critical reflection by clearly engaging preservice teachers in questioning authoritative practices embedded in current practice and considering issues of social justice, equity, and meaning. Glenda Moss (2008) investigated this dilemma and suggested such a possibility. She stated,

Can preservice teachers enter the teaching profession with wisdom in their ‘bag of tricks’? This concept aligns with Dewey’s (1910) view of reflection as a kind of interruption in tacit knowledge within the stream of practice that results in changed action. That is the goal of moving portfolio assessment to the point of preservice teachers’ developing a critical lens through which they can examine the knowledge banks that they are being provided and so eventually develop teaching stances that are ideally grounding in rigorous inquiry and visions beyond their students’ passing standardized tests. (p. 155)

Pamela Moss (1998) proposed a portfolio assessment using the framework of philosophical hermeneutics. Moss described the possibility of a locally-scored system of portfolio evaluation using evaluators that have been chosen because they know the candidates well. Her vision included teams of raters who compared their evaluations and collaborated to offer feedback, providing data triangulation rooted in the specific context. Moss (1998) called for further research towards an “integrative approach” to portfolio assessment rather than one that was strictly parametric but noted, “However, existing theory in educational measurement does not provide an adequate epistemological basis to support this promising work” (p. 205). Delandshere and Arens (2003) echoed Moss’s (1998) concerns when they stated, “The standards-based movement and its existing or

proposed high-stakes assessment and accreditation mandates are still working from the measurement-driven logic” (p. 71).

Visions of teacher assessment that gaze beyond standardization toward meaning and *phronesis* (wisdom) entail shifting conceptions of validity (Moss, 1998). The very act of trying to force portfolios into a parametric paradigm may be antithetical to the deeper meaning of reflection (Meeus, Petegem, & Engels, 2009). Yet rigorous standards for responsible research prevent teacher educators from ignoring empirical concerns for validity and reliability. Further research may reveal the means to strike a much-needed balance. Perhaps a clearly written rubric, such as the *REPORT* created for this study, is one step down the path of the integrative approach called for by Moss (1998) and Delanshere and Arens (2003).

### **Summary of Study Discussion**

The need for highly effective teachers for all students is clear in the research literature. Effective teachers are needed who are able to reflect on their practice in order to improve their own teaching and enhance student learning. It is the responsibility of teacher educators to prepare such teachers for the profession. The results of this study support the conclusion that portfolios comprise a valid and reliable assessment tool to promote high levels of reflection and increase teachers’ capacity for excellent performance.

**Implications for practice.** The study findings and interpretations suggest a variety of implications for practice. First, teacher education personnel can create valid and reliable rubrics for portfolios. Second, in-depth reflection can be developed using training and prompts. Finally, highly reflective portfolios require substantial time and effort to



produce, implying a great investment of time and resources by teacher preparation programs wishing to use them.

**Future research was recommended in four areas.** Findings and interpretations from this study point to several recommendations for future research. The first recommendation is to design large-scale studies to investigate increased methods of valid and reliable portfolio assessments. A second recommendation is to design and conduct studies that clarify excellent teacher performance, including constructs such as teacher reflective capability displayed in portfolios. Third, this study suggests that further research is needed that links portfolio assessment to K-12 student learning. And finally, studies of qualitative and hermeneutic methods of portfolio assessment not rooted in quantitative standards for reliability and validity may benefit teacher education programs and help ease the tension between formative and summative aspects of portfolio evaluation.

In conclusion, the results from this study support the notion that portfolios can validly and reliably assess preservice teacher reflective capability, given that sufficient training and support are provided to both portfolio creators and to portfolio assessors. Such training takes time and effort, but results of this study indicate that it can contribute to the development of higher levels of reflection in preservice teachers. Even with extensive instruction and support, some preservice teachers still find in-depth reflective writing to be challenging. While optimism is in order that deep reflection will both enhance teacher performance and increase K-12 student achievement, further research is needed to substantiate such claims. Since teacher preparation programs constitute unique contexts, each institution would do well to conduct its own cost-benefit analysis to

determine the relative value of its investment of time in standards-based portfolios for evaluating preservice teacher reflection.

## References

- Adams, T. (1995). A paradigm for portfolio assessment in teacher education. *Education, 115*(4), 568-572.
- Amobi, F. (2006). Beyond the call: Preserving reflection in the preparation of “highly-qualified” teachers. *Teacher Education, 33*(2), 23-35.
- Anderson, R., & DeMeulle, L. (1998). Portfolio use in twenty-four teacher education programs. *Teacher Education Quarterly, 25*, 23-31.
- Barone, D., & Morrell, E. (2007). Multiple perspectives on preparing teachers to teach reading. *Reading Research Quarterly, 42*(1), 167-180.
- Barrett, H. (2007). Researching electronic portfolios and learner engagement: The REFLECT initiative. *Journal of Adolescent & Adult Literacy, 50*(6), 436-439.
- Barrett, H., & Wilkerson, J. (2004). *Conflicting paradigms in electronic portfolio approaches: Choosing an electronic portfolio strategy that matches your conceptual framework*. Retrieved from <http://www.helenbarrett.com/systems/paradigms.html>
- Bartlett, A. (2006). It was hard work but it was worth it: ePortfolios in teacher education. In A. Jafari & C. Kaufman (Eds.), *Handbook of research on ePortfolios* (pp. 327-339). Hershey, PA: Idea Group Reference.
- Barton, J., & Collins, A. (1993). Portfolios in teacher education. *Journal of Teacher Education, 44*(3), 200-210.
- Bauer, W., & Dunn, R. (2003). Digital reflection: The electronic portfolio in music education. *Journal of Music Teacher Education, 13*(1), 7-20.

- Bird, T. (1990). The schoolteacher's portfolio: An essay on possibilities. In J. Millman and L. Darling-Hammond (Eds.), *The new teacher handbook of teacher evaluation: Assessing elementary and secondary school teachers* (pp. 241-256). Newbury Park, CA: Corwin Press, Inc.
- Boody, R. (2008). Teacher reflection as teacher change, and teacher change as moral response. *Education*, 128, 498-506.
- Borko, H., Michalec, P., Timmons, M., & Siddle, J. (1997). Student teaching portfolios: A tool for promoting reflective practice. *Journal of Teacher Education*, 48(5), 345-358.
- Brookhart, S., & Nitko, A. (2008). *Assessment and grading in classrooms*. Upper Saddle River, NJ: Pearson.
- Burns, M., & Haight, S. (2005). Psychometric properties and instructional utility of assessing special education teacher candidate knowledge with portfolios. *Teacher Education and Special Education*, 28(314), 185-194.
- Carney, J. (2004, April). *Setting an agenda for electronic portfolio research: A framework for evaluating portfolio literature*. Paper presented at the annual meeting of the American Educational Research Association, San Diego, CA. Retrieved from <http://it.wce.wvu.edu/carney/Presentations/AERA04/AERAresearchlit.pdf>
- Carney, J. (2006). Analyzing research on teachers' electronic portfolios: What does it tell us about portfolios and methods for studying them? *Journal of Computing in Teacher Education*, 22(3), 89-97.

- Cawyer, C., & Caldwell, D. (2002). Performance-based assessment of teacher education: Some contextual considerations. *Action in Teacher Education*, 23(4), 1-3.
- Covenant College. (2009). *Student teaching handbook, fall 2009*. Lookout Mountain, GA: Covenant College.
- Darling-Hammond, L., & Youngs, P. (2002). Defining highly-qualified teachers: What does scientifically-based research actually tell us? *Educational Researcher*, 31(9), 13-25.
- Delandshere, G., & Arens, S. (2003). Examining the quality of evidence in preservice teacher portfolios. *Journal of Teacher Education*, 54(1), 57-73.
- Derham, C., & Diperna, J. (2007). Digital professional portfolios of preservice teaching: An initial study of score reliability and validity. *Journal of Technology and Teacher Education*, 15(3), 363-381.
- Dewey, J. (1933). *How we think: A restatement of the relation of reflective thinking to the educative process*. Boston: Houghton-Mifflin.
- Dhonau, S., & McAlpine, D. (2005). An electronic portfolio for the ACTFL/NCATE teacher program standards in the Second Language Methods Course 1. *Foreign Language Annals*, 38(1), 69-77.
- Ducharme, E., & Ducharme, M. (1996). Reflecting on reflecting. *Journal of Teacher Education*, 47(2), 83-84. doi: 10.1177/0022487196047002001
- Evans, S., Daniel, T., Mikovch, A., Metze, L., & Norman, A. (2006). The use of technology in portfolio assessment of teacher education. *Journal of Technology and Teacher Education*, 14(1), 5-27.

- Gathercoal, P., Love, D., & McKean, G. (2007, April). *California Lutheran University's School of Education webfolios in teacher education: Teacher performance expectations (TPE's) and teaching performance assessments (TPA's)-present and future*. Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL.
- Gatlin, L., & Jacob, S. (2002). Standards-based digital portfolios: A component of authentic assessment for preservice teachers. *Action in Teacher Education*, 23(4), 35-42.
- Gay, L., Mills, G., & Airasian, P. (2006). *Educational research: Competencies for analysis and application* (8<sup>th</sup> ed.). Upper Saddle River, NJ: Pearson.
- Goldsby, D., & Fazal, M. (2001). Now that your students have created web-based digital portfolio, how do you evaluate them? *Journal of Technology and Teacher Education*, 9, 607-616.
- Good, J., & Weaver, A. (2003). Evaluation of preservice teacher internships: A model to encourage career continuity and program reform. *Journal of Personnel in Education*, 17(3), 263-275.
- Gordinier, C., Conway, K., & Journet, A. (2006). Facilitating teacher candidates' reflective development through the use of portfolios, teacher work sample, and guided reflections. *Teaching & Learning*, 20(2), 89-105.
- Hallman, H. (2007). Negotiating teacher identity: Exploring the use of electronic teaching portfolios with preservice English teachers. *Journal of Adolescent & Adult Literacy*, 50(6), 474-485.

- Hartmann, C. (2004). Using teacher portfolios to enrich the methods course experiences of prospective mathematics teachers. *School Science and Mathematics, 104*(8), 392-406.
- Hatton, N., & Smith, D. (1995). Reflection in teacher education: Towards definition and implementation. *Teaching & Teacher Education, 11*(1), 33-49.
- Herman, J., & Winters, L. (1994). Portfolio research: A slim collection. *Educational Leadership, 52*(2), 48-55.
- Hicks, T., Russo, A., Autrey, T., Gardner, R., Kabodian, A., & Edington, C. (2007). Rethinking the purposes and processes for designing digital portfolios. *Journal of Adolescent & Adult Literacy, 50*(6), 450-458.
- Hill, D. (2003). E-Folio and teacher candidate development. *The Teacher Educator, 38*(4), 256-266.
- Hinkle, D., Wiersma, W., & Jurs, S. (2003). *Applied statistics for the behavioral sciences* (5<sup>th</sup> ed.). Boston: Houghton Mifflin.
- Horton, M. (2004). DIGITAL portfolios in physical education teacher preparation. *Journal of Physical Education, Recreation, and Dance, 75*(9), 35-39.
- Johnson, C. (2006). The analytic assessment of online portfolio in undergraduate technical communication: A model. *Journal of Engineering Education, 95*(4), 279-287.
- Johnson, R., Mims-Cox, J., & Doyle-Nichols, A. (2006). *Developing portfolios in education: A guide to reflection, inquiry, and assessment*. Thousand Oaks, CA: SAGE Publications.

- King, F., & Roblyer, M. (1984). Alternative designs for evaluating computer-based instruction. *Journal of Instructional Development*, 7(3), 23-29.
- Klecker, B. (2000). Content validity of preservice teacher portfolios in a standards-based program. *Journal of Instructional Psychology*, 27(1), 35-38.
- Klenowski, V. (2000). Portfolios: Promoting teaching. *Assessment in Education*, 7(2), 215-238.
- Ledoux, M., & McHenry, N. (2006). Electronic portfolio adoption for teacher education candidates. *Early Childhood Education Journal*, 34(2), 103-116.
- Lee, H. (2005). Understanding and assessing preservice teachers' reflective thinking. *Teaching and Teacher Education*, 21, 699-715. doi: 10.1016/j.tate.2005.05.007
- Lee, I. (2008). Fostering preservice reflection through response journals. *Teacher Education Quarterly*, 35(1), 117-139.
- Lee, M., & Hare, M. (2007). A web-based learning portfolio system. *Journal of Physical Education, Recreation, and Dance*, 78(9), 45-49.
- Levitt, K., & Schreiber, J. (2008, March). *Using teacher candidate portfolios to demonstrate impact on K-6 student learning*. Paper presented at the annual meeting of the American Educational Research Association, New York, NY.
- Loughran, J. (2002). Effective reflective practice: In search of meaning in learning about teaching. *Journal of Teacher Education*, 53(1), 33-43.
- Loughran, J., & Corrigan, D. (1995). Teaching portfolios: A strategy for developing learning and teaching in preservice education. *Teaching & Teacher Education*, 11, 565-577.



- Lynch, L., & Purnawarman, P. (2004). Electronic portfolio assessments in U.S. educational and instructional technology programs: Are they supporting teacher education? *TechTrends*, 48(1), 50-57.
- Lyons, N. (1998). Constructing narratives for understanding: Using portfolio interviews to scaffold teacher reflection. In Lyons, N. (Ed.), *With portfolio in hand: Validating the new teacher professionalism* (pp. 23-37). New York: Teachers College Press.
- Mansvelder-Longayroux, D., Beijaard, D., Verloop, N., & Vermunt, J. (2007). Functions of the learning portfolio in student teachers' learning process. *Teachers College Record*, 109(1), 126-159.
- McCormick, T., Sunal, C., & Sunal, D. (2005). The demonstration of teaching values in elementary pre-service teachers' e-portfolios. *International Journal of Social Education*, 20(1), 81-95.
- McFarland, J., Wisniewski, S., & Vermette, P. (1997, October). *Comparative ratings of the utility of portfolio requirements: Toward content validity*. Paper presented at the Annual Meeting of the Northeastern Educational Research Association, Ellenville, NY. Retrieved from ERIC database. (ED414339)
- Meeus, W., Petegem, P., & Engels, N. (2009). Validity and reliability of portfolio assessment in pre-service teacher education. *Assessment & Evaluation in Higher Education*, 34, 401-403. doi: 10.1080/02602930802062659
- Meyer, D., & Tusin, L. (1999). Preservice teacher's perceptions of portfolios: Process versus product. *Journal of Teacher Education*, 50(2), 131-139.

- Milman, N. (2005). Web-based digital teaching portfolios: Fostering reflection and technology competence in preservice teacher education students. *Journal of Technology and Teacher Education, 13*(3), 373-396.
- Morgan, B. (1999). Portfolios in a preservice teacher field-based program: Evolution of a rubric for performance assessment. *Education, 119*(3), 416-427.
- Moss, G. (2008). Developing a critical lens among preservice teachers while working within mandated performance-based assessment systems. *Teacher Education and Practice, 21*, 146-159.
- Moss, P. (1998). Rethinking validity for the assessment of teaching. In Lyons, N. (Ed.), *With portfolio in hand: Validating the new teacher professionalism* (pp. 202-219). New York: Teachers College Press.
- Naizer, G. (1997). Validity and reliability issues of performance-portfolio assessment. *Action in Teacher Education, 18*(4), 1-9.
- Ntuli, E., Keengwe, J., & Kyei-Blankson, L. (2009). Electronic portfolios in teacher education: A case study of early childhood teacher candidates. *Early Childhood Education Journal, 37*, 121-126. doi: 10.1007/s10643-009-0327-y
- Orland-Barak, L. (2005). Portfolio as evidence of reflective practice: What remains 'untold.' *Educational Research, 47*(1), 25-44.  
doi:10.1080/0013188042000337541
- Pecheone, R., Pigg, M., Chung, R., & Souviney, R. (2005). Performance assessment and electronic portfolios: Their effect on teacher learning and education. *The Clearing House, 78*(40), 164-177.

- Pennington, R. (2009, April). *Preservice teacher portfolios: How do they mediate teacher reflection?* Paper presented at the annual meeting of the American Educational Research Association, San Diego, CA.
- Penny, C., & Kinslow, J. (2006). Faculty perceptions of electronic portfolios in a teacher education program. *Contemporary Issues in Technology and Teacher Education*, 6(4), 418-435.
- Popham, W. J. (2006). *Assessment for educational leaders*. Boston: Pearson.
- Reidinger, B. (2006). Mining for meaning: Teaching students how to reflect. In A. Jafari & C. Kaufman (Eds.), *Handbook of research on ePortfolio* (pp. 91-101). Hershey, PA: Idea Group Reference.
- Reis, N., & Villaume, S. (2002). The benefits, tensions, and visions of portfolios as a wide-scale assessment for teacher education. *Action in Teacher Education*, 23(4), 10-17.
- Rickards, W., Diez, M., Ehley, L., Guilbault, L., Locker, G., Hart, J., & Smith, P. (2008). Learning, reflection, and electronic portfolios: Stepping toward an assessment practice. *The Journal of General Education*, 57(1), 31-50. Retrieved from [http://www.psupress.psu.edu/journals/jnls\\_jge.html](http://www.psupress.psu.edu/journals/jnls_jge.html)
- Ring, G., & Foti, S. (2003, March/April). Addressing standards at the program level with electronic portfolios. *TechTrends*, 47, 28-32.
- Ring, G., & Foti, S. (2006). Using ePortfolios to facilitate professional development among pre-service teachers. In A. Jafari & C. Kaufman (Eds.), *Handbook of research on ePortfolio* (pp. 340-355). Hershey, PA: Idea Group Reference.

- Robinson, L., & Kelley, B. (2007). Developing reflective thought in preservice educators: Utilizing role-plays in digital video. *Journal of Special Education Technology*, 22(2), 32-43.
- Rodgers, C. (2002). Defining reflection: Another look at John Dewey and reflective thinking. *Teachers College Record*, 104, 842-866.
- Schön, D. (1983). *The reflective practitioner: How professionals think in action*. Cambridge, MA: Basic Books, Inc.
- Schön, D. (1987). *Educating the reflective practitioner: Toward a new design for teaching and learning in the professions*. San Francisco, CA: Jossey-Bass.
- Schwartz, S. & Rolheiser, C. (2001). Pre-service portfolios: A base for professional growth. *Canadian Journal of Education*, 26(3).
- Shapley, K., & Bush, M. (1999). Developing a valid and reliable portfolio assessment in the primary grades: Building on practical experience. *Applied Measurement in Education*, 12(2), 111-132.
- Shulman, L. (1998). Teacher portfolios: A theoretical activity. In Lyons, N. (Ed.), *With portfolio in hand: Validating the new teacher professionalism* (pp. 23-37). New York: Teachers College Press.
- Spalding, E., & Wilson, A. (2002). Demystifying reflection: A study of pedagogical strategies that encourage reflective journal writing. *Teachers College Record*, 104, 1393-1421.
- Sparks-Langer, G., & Colton, A. (1991). Synthesis of research on teacher's reflective thinking. *Educational Leadership*, 48(6), 37-44.

- Sparks-Langer, G., Simmons, J., Pasch, M., Colton, A., & Starko, A. (1990). Reflective pedagogical thinking: How can we promote it and measure it? *Journal of Teacher Education, 41*(4), 23-32. doi: 10.1177/002248719004100504
- Stronge, J., Ward, T., Tucker, P., & Hindman, J. (2007). What is the relationship between teacher quality and student achievement? An exploratory study. *Journal of Personnel Evaluation in Education, 20*, 165-184. doi: 10.1007/s11092-008-9053-z
- Strudler, N., & Wetzel, K. (2005). The diffusion of electronic portfolios in teacher education: Issues of initiation and implementation. *Journal of Research on Technology in Education, 37*(4), 411-434.
- Strudler, N., & Wetzel, K. (2008). Costs and benefits of electronic portfolios in teacher education: Faculty perspectives. *Journal of Computing in Teacher Education, 24*(4), 135-142.
- Sulzen, J. (2007). *Identifying judgments supported by preservice teacher electronic portfolios*. Unpublished doctoral dissertation, University of Connecticut, Storrs.
- Sulzen, J., & Young, M. (2007, April). *Fast and reliable evaluation of preservice teacher electronic portfolios*. Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL.
- Thompson, A. (2005). Scientifically based research: Establishing a research agenda for technology in the teacher education community. *Journal of Research on Technology in Education, 37*(4), 331-337.

- Tigelaar, D., Dolmans, D., Wolfhagen, I., & van der Vleuten, C. (2005). Quality issues in judging portfolios: Implications for organizing teaching portfolio assessment procedures. *Studies in Higher Education, 30*, 595-610.
- Tillema, H., & Smith, K. (2007). Portfolio appraisal: In search of criteria. *Teaching and Teacher Education, 23*, 442-456. doi: 10.1016/j.tate.2006.12.005
- Valli, L. (1997). Listening to other voices: A description of teacher reflection in the United States. *Peabody Journal of Education, 72*(1), 67-88. Retrieved from [http://vnweb.hwwilsonweb.com/hww/results/results\\_single\\_ftPES.jhtml](http://vnweb.hwwilsonweb.com/hww/results/results_single_ftPES.jhtml)
- Van Manen, M. (1977). Linking ways of knowing with ways of being practical. *Curriculum Inquiry, 6*(3), 205-228.
- Van Sickle, M., Bogan, M., & Kamen, M. (2005). Dilemmas faced establishing portfolio assessment of pre-service teachers in the southeastern United States. *College Student Journal, 39*(3), 497-509.
- Watts, M., & Lawson, M. (2009). Using a meta-analysis activity to make critical reflection explicit in teacher education. *Teaching and Teacher Education, 25*, 609-616.
- Welsch, R., & Devlin, P. (2006). Developing preservice teacher's reflection: Examining the use of video. *Action in Teacher Education, 28*(4), 53-61.
- Wenzlaff, T., & Cummings, K. (1996, Winter). The portfolio as metaphor for teacher reflection. *Contemporary Education, 67*, 109-112.
- Wetzel, K., & Strudler, N. (2005). The diffusion of electronic portfolios in teacher education: Next steps and recommendations from accomplished users. *Journal of Research on Technology in Education, 38*(2), 231-244.

- Wetzel, K., & Strudler, N. (2006). Costs and benefits of electronic portfolios in teacher education: Student voices. *Journal of Computing in Teacher Education*, 22(3), 99-108.
- Wilhelm, L., Puckett, K., Beisser, S., Wishart, W., Merideth, E., & Sivakumaran, T. (2006). Lessons learned from the implementation of electronic portfolios at three universities. *TechTrends*, 50(4), 62-65.
- Winsor, P., Butt, R., & Reeves, H. (1999). Portraying professional development in preservice teacher education: Can portfolios do the job? *Teachers and Teaching: Theory and Practice*, 5(1), 9-13.
- Wolf, K. (1991). The schoolteacher's portfolio: Issues in design, implementation, and evaluation. *Phi Delta Kappan*, 73, 129-136.
- Woodward, H., & Nanlohy, P. (2004). Digital portfolios in pre-service teacher education. *Assessment in Education*, 11(2), 167-178.
- Wray, S. (2007). Teaching portfolios, community, and pre-service teachers' professional development. *Teaching and Teacher Education*, 23, 1139-1152. doi: 10.1016/j.tate.2006.10.004
- Wray, S. (2008). Swimming upstream: Shifting the purpose of an existing teaching portfolio requirement. *The Professional Educator*, 32(1). Retrieved from [http://www.auburn.edu/academic/societies/professional\\_educator/articles/wray\\_final.pdf](http://www.auburn.edu/academic/societies/professional_educator/articles/wray_final.pdf)
- Yao, Y., Aldrich, J., & Foster, K. (2008, March). *Preservice teachers' perceptions of an electronic portfolio as a tool for reflection and teacher certification*. Paper

presented at the annual meeting of the American Educational Research Association, New York, NY.

Yao, Y., Thomas, M., Nickens, N., Downing, J., Burkett, R., & Lamson, S. (2008).

Validity evidence of an electronic portfolio for preservice teachers. *Educational Measurement: Issues and Practice*, 27(1), 10-24.

Zeichner, K., & Wray, S. (2001). The teaching portfolio in U.S. teacher education programs: What we know and what we need to know. *Teaching and Teacher Education*, 17, 613-621.

Zellers, M., & Mudrey, R. (2007). Electronic portfolios and metacognition: A phenomenological examination of the implementation of e-portfolios from the instructor's perspective. *International Journal of Instructional Media*, 34(4), 419-430.



## **Appendices**

## Appendix A

### Rubric for Evaluating Portfolio Reflective Thinking (REPORT)

Type of Reflection	Level 0	Level 1	Level 2	Level 3
Technical/Descriptive	<p><b>(0)</b> Lists artifact and states artifact topic or skill only OR restates the standard</p>	<p><b>(1)</b> Reports the event or experience that forms the artifact content ; basic description of content of artifact; may include statement of reason without explanation (Orland-Barak, 2005)</p> <p><i>“This was a two week unit for science class. The unit was on the solar system, the planets, and the moon.”</i> <i>“The reason this unit was chosen was in part because I wanted to incorporate as many disciplines as was possible.”</i></p>	<p><b>(2)</b> Describes artifact AND explains reasons for artifact content based on external criteria (standards, “best practice”) or general principles; applies theory to practice in light of own experience only</p> <p><i>“I felt this science experiment was beneficial in showing the students how their sense of taste works with their sense of smell. I feel it is important to allow students to see that things need other things to work, just like people need other people.”</i></p>	<p><b>(3)</b> Describes artifact AND explains reasons for artifact content based on specific principles or theory; cites evidence from the artifact directly to show application of theory to practice and connections to standards.</p> <p><i>I have included in my portfolio two classroom observations of children at play to demonstrate my understanding of how children learn through interactions with others. The constructivist theory believes children should actively construct knowledge and explore their world together. I observed children setting boundaries and preferences, communicating verbally and nonverbally, and how they responded to teacher and student interactions. This play time gave children an opportunity to learn, build motor skills, and relationships. The observations are reminders to me that children can</i></p>

				<i>learn in collaborative settings and can benefit from a variety of learning experiences.</i>
Personal Growth	<p><b>(0)</b> Does not relate artifact to personal growth, beliefs, feelings or values at all.</p>	<p><b>(1)</b> Expresses feelings or beliefs about what constitutes good teaching; explains the value or importance of the standard but with little reference to the artifact. (Valli, 1997)</p> <p><i>“It is important for teachers to have strong colleague, parent, and community connections. Having these strong connections only enhances the students’ learning.”</i></p> <p><i>“While teachers cannot physically observe all student interactions if they model Christ-like words and behavior, they can be change agents in future ways their students work and play together.”</i></p>	<p><b>(2)</b> Expresses growth from experience represented in artifact by stating <i>that</i> something was learned without specific evidence from the artifact to exemplify this learning.</p> <p><i>“I wanted to put these two artifacts in my portfolio because I think they represent my growth in using technology.”</i></p>	<p><b>(3)</b> Expresses growth from experience represented in artifact; cites evidence from artifact for growth and offers suggestions for improved practice OR Expresses growth across time using evidence from multiple artifacts.</p> <p><i>“I learned one good lesson from this lesson. Before creating the words, I handed out the different letters to the students to hold while they waited their turn to stand up and insert thier [sic] letter sound to help create the word. However, there was a lot of rustling with the paper plates while students were waiting to go up. If I were to do this lesson again, (which I plan on doing, just with another word family) I will hold all of the extra plates and select those students who are sitting properly and quietly to stand up and help create a word.</i></p>

Dialogic/Critical	<p><b>(0)</b> Does not discuss artifact's impact on others at all, so multiple viewpoints and impact on ethical, moral and justice issues is not included.</p>	<p><b>(3)</b> Explains how work represented in artifact impacts others (student learning, peers, parents, administrators);</p> <p><i>"I then administered, scored, and analyzed the post tests. I am pleased to say that I see progress in what my students know. I also realize that, if I were to teach the unit again, should have been emphasized even more. Sequencing events is something that almost every student missed on both exams."</i></p>	<p><b>(4)</b> Weighs competing claims and multiple viewpoints as one analyzes artifacts; explains alternative solutions to a problems that may have been encountered in teaching situation represented in artifact</p> <p><i>"This DIBELS score shows that this student is at risk for nonsense word fluency and needs to have intervention. But she is reading on a first grade level fluently so she can obviously read. I think we need to use various assessment tools together to determine whether a child needs intervention."</i></p>	<p><b>(5)</b> Questions practices of the teaching profession represented in artifact ("best-practice", standards, testing, etc.) based on ethical, moral, or justice concerns</p> <p><i>"This unit includes a variety of researched-based reading strategies, but not much social studies content. In fact, during student teaching my cooperating teacher didn't teach social studies at all. It seems that if kids are going to learn to be productive, democratic citizens, they need to have knowledge of history and government. The kids that don't have as many privileges and experience need that knowledge to succeed on tests and in life. I think not teaching content like social studies just makes the 'achievement gap' wider."</i></p>
-------------------	--	--	---	---

*Rubric for Evaluating Portfolio Reflective Thinking (REPORT) -Score Sheet*

Name/Number: \_\_\_\_\_

Domain A: Knowledge	Type of Reflection	Level 0	Level 1	Level 2	Level 3
	Technical/Descriptive	<b>0</b>	<b>1</b>	<b>2</b>	<b>3</b>
	Personal Growth	<b>0</b>	<b>1</b>	<b>2</b>	<b>3</b>
	Dialogic/Critical	<b>0</b>	<b>1</b>	<b>2</b>	<b>3</b>

Domain score: \_\_\_\_\_

Domain B: Planning	Type of Reflection	Level 0	Level 1	Level 2	Level 3
	Technical/Descriptive	<b>0</b>	<b>1</b>	<b>2</b>	<b>3</b>
	Personal Growth	<b>0</b>	<b>1</b>	<b>2</b>	<b>3</b>
	Dialogic/Critical	<b>0</b>	<b>1</b>	<b>2</b>	<b>3</b>

Domain score: \_\_\_\_\_

Domain C: Instruction	Type of Reflection	Level 0	Level 1	Level 2	Level 3
	Technical/Descriptive	<b>0</b>	<b>1</b>	<b>2</b>	<b>3</b>
	Personal Growth	<b>0</b>	<b>1</b>	<b>2</b>	<b>3</b>
	Dialogic/Critical	<b>0</b>	<b>1</b>	<b>2</b>	<b>3</b>

Domain score: \_\_\_\_\_

Domain D: Assessment	Type of Reflection	Level 0	Level 1	Level 2	Level 3
	Technical/Descriptive	<b>0</b>	<b>1</b>	<b>2</b>	<b>3</b>
	Personal Growth	<b>0</b>	<b>1</b>	<b>2</b>	<b>3</b>
	Dialogic/Critical	<b>0</b>	<b>1</b>	<b>2</b>	<b>3</b>

Domain score: \_\_\_\_\_

Domain E: Classroom Environment	Type of Reflection	Level 0	Level 1	Level 2	Level 3
	Technical/Descriptive	<b>0</b>	<b>1</b>	<b>2</b>	<b>3</b>
	Personal Growth	<b>0</b>	<b>1</b>	<b>2</b>	<b>3</b>
	Dialogic/Critical	<b>0</b>	<b>1</b>	<b>2</b>	<b>3</b>

Domain score: \_\_\_\_\_

Domain F: Professional Growth	Type of Reflection	Level 0	Level 1	Level 2	Level 3
	Technical/Descriptive	<b>0</b>	<b>1</b>	<b>2</b>	<b>3</b>
	Personal Growth	<b>0</b>	<b>1</b>	<b>2</b>	<b>3</b>
	Dialogic/Critical	<b>0</b>	<b>1</b>	<b>2</b>	<b>3</b>

Total Score: \_\_\_\_\_ / Domain score: \_\_\_\_\_

## Appendix B

### *Portfolio Reflective Writing Guide*

#### **Introduction**

Beginning with Dewey's (1933) concept of reflection as rational problem-solving, teacher educators have considered reflective thinking essential to improving practice. Schön's (1983, 1987) work increased focus on reflection as a way for teachers to frame and solve problems within the complex context of teaching situations (Loughran, 2002). By careful reflection on experience over time, teachers develop professional knowledge and connect theory to practice (Lee, 2008; Loughran, 2002; Van Manen, 1977). In essence, effective reflection leads to effective teaching (Loughran, 2002).

Reflection has not been precisely defined and attempts to measure it have produced ambiguous results (Rodgers, 2002). Van Manen (1977) offered one of the first taxonomies for describing reflection. Van Manen proposed three levels of reflectivity: technical-rational, deliberative, and critical (Boody, 2008). Technical-rational reflectivity is concerned with determining how efficiently methods and means accomplish the predetermined ends or objectives of instruction. In other words, how effectively has the teaching method achieved the goals set for it by theory or outside authority?

Van Manen's (1977) second level of reflectivity (deliberative) calls for "an interpretive understanding both of the nature and quality of educational experience" (p. 226). In this level, teachers recognize their own value commitments to a particular interpretive framework as they make judgments about education practices (curriculum, methods, etc.).

Finally, Van Manen (1977) proposes a higher level of reflectivity (critical reflection) aimed at pondering "worthwhile educational ends" on the basis of "justice, equality, and freedom" (p. 227). In this critical level, teachers consider the political, moral, and ethical impact of established educational practices.

Specifically, researchers claim that preservice teacher portfolios enhance reflection. The Teacher Education Developmental Portfolio offers one opportunity to reflect on your work and demonstrate your ability to think reflectively by writing reflective statements to accompany each artifact and reflective essays that describe your growth over time as you have completed the teacher education program. This *Portfolio Reflective Writing Guide* offers a rubric, prompts, and questions designed to guide you to think deeply about all aspects of your teaching and learning progress to this point.

## Guiding Questions for Reflection

1. Read this introduction and the accompanying *Rubric for Evaluating Portfolio Reflective Thinking (REPORT)*.
2. Think of a particular artifact that represents a significant learning experience that helped you grow or a learning problem that you overcame.
3. Use the following questions to analyze your thoughts and write a reflective statement:
  - a. Technical/Rational:
    - i. What was the main problem or key learning point that I took away from this assignment or experience?
    - ii. What educational theories or ideas did I need to apply when I experienced this event?
    - iii. What strengths and areas for improvement did this experience or event reveal to me about my own performance?
    - iv. What might I do differently the next time I did a similar activity or event?
    - v. How does this experience or artifact demonstrate that I am capable of meeting the teacher standards expected of me?
    - vi. What specific evidence from the artifact or experience serves as evidence of my capability?
  - b. Personal Growth
    - i. What strengths and areas for improvement did this experience or event reveal to me about my own performance?
    - ii. When writing the reflective essays, how can I describe my growth over time?
    - iii. What can I do well now that I could not do when I began this program?
    - iv. What specific evidence from either one artifact or multiple artifacts shows what I have learned both in single events and over time?
  - c. Dialogic/Critical
    - i. How have I applied my own values, moral judgments, and ethical commitments as I have experienced teaching?
    - ii. How has my teaching or educational decision-making influenced others:
      1. Student learning?
      2. My colleagues at school?
      3. Parents?

4. The larger community?
  - iii. Are there multiple ways to solve any problems posed in this artifact or event? If so, did I consider them?
  - iv. What might I have done differently in this situation?
  - v. What, if anything, concerns me about how students are treated?
  - vi. Are the “best practices” I am expected to accomplish promoting fairness, justice, and high levels of ethical and moral standards according to my own values?

### **Discussion and Feedback**

1. Select one artifact from your current portfolio or from the assignments and activities you have completed during this program.
2. Write a reflective statement for this using your textbook, the information in the *REPORT*, and the guiding questions above.
3. Share your reflective statement with a partner. Use the questions to give each other feedback and suggestions for revision.
4. Revise your reflective statement and type it into your portfolio. Submit to the professor for further feedback at a later date.



Vita  
**Rebecca E. Pennington**

**Personal**

**Family**

- Married to Aubrey Tucker Pennington, Jr.

**Children:**

- Hannah Pennington and Aubrianna Pennington

**Experience**

**July 1, 2002 to Present (2010) Covenant College Lookout Mountain, GA**

**Assistant Professor of Education**

**Courses Taught:**

- EDU 315-316 Teaching Reading in the Early and Middle Grades
- EDU 313-314 Language Arts Content and Skills in the Early and Middle Grades
- EDU 410-1 Educational Assessment
- Student Teaching Supervising
- EDU 366/377 Literature for Children in the Early Grades/Middle Grades
- EDU 328 Social Studies Content and Skills in the Early Grades
- EDU 394 Senior Integration Paper
- COR 100 Christian Mind
- EDU 317 Assessing and Correcting Reading Difficulties
- EDU 318 Reading and Writing in the Content Areas
- Variety of courses in Bachelor of Science in Early Childhood Education Program through Covenant's adult degree completion program

**October 2001-June 2002 Manatee Middle School Naples, FL**  
**ESOL Immersion Teacher**

**August 2000-October 2001 Naples Christian Academy Naples, FL**  
**Curriculum Specialist**

**1998-2000 Manatee Elementary School Naples, FL**  
**Third Grade Teacher**  
Florida State Certification (Professional #646272)-Elementary, Through 2006

**1986-1998 Naples Christian Academy Naples, FL**  
**Elementary Grade Teacher/Middle School Teacher/Curriculum Coordinator**

**1982-1986 Chickamauga Elementary School Chickamauga, GA**  
**Elementary Teacher**  
Fifth grade teacher, Georgia State Certification

## Education

### Covenant College

Lookout Mountain, GA

B.A., Elementary Education, *cum laude*, 1982

Master of Education, Integrated Curriculum and Instruction, 2002

Graduate school, University of Tennessee at Chattanooga, Ed. D. in Learning and Leadership, ABD

## Professional Service

Trained Peer Evaluator, Florida POP; Mentor Teacher 1994-1998

ACSI Accreditation Team Member for First Academy in Orlando, FL

Presented Curriculum Workshops at the Florida State ACSI Conference,  
November 1995, 1997

Presented Seminar: "Integrating the Curriculum: What, Why and How"-Manatee  
Middle School, Naples, FL, June 2002

Member: School Council, 2003-2005, Chattanooga Valley Elementary School

Member: Spelling Curriculum Revision Team (consultant) for Association of Christian  
Schools International, January 30-31, 2006, Colorado Springs, CO.

#### Professional Associations:

American Educational Research Association, 2007-2009

Association for Supervision and Curriculum Development, 2002-2009

International Reading Association, 2003-2009

#### College Committees:

Assessment Committee, Education Department

Georgia State Professional Standards Commission Self-study Committee

Library Committee

Chapel Committee-chair 2005-2006

Strategic Planning Working Committee-Fall 2008

#### Graduate Assistant:

Teaching Assistant to Dr. Philip Horton in EDU 620, Introduction to Research,  
Covenant College Master of Education Program (Summer 2003-2005)

#### Conferences:

AERA Annual Meeting, March 2008, New York: Discussant for SIG Paper  
Session on Education and the World Wide Web

AERA Annual Meeting, April 2009, San Diego, Presented Roundtable  
paper, Preservice Teacher Portfolios: How Do They Mediate Teacher  
Reflection?

#### Presentations:

Workshop with Dr. Donovan Graham at the Christian Schools International  
Leadership Convention: Pre-Convention Workshop entitled, "Does Grace Have a  
Place in Your School?", July 21-22, 2003

Participated in faculty workshop funded by Lily entitled, "Imagining the Arts at

Covenant", August 5-8, 2003

February 13, 2004 and January 2006: Seminar presentations at ACSI Convention in Birmingham,

"Beyond the Reading Wars: Are we still fighting?"

GATE Conference, Atlanta, Georgia, October 2004: Panel presentation,

"Paired and Prepared: Preservice Reading Teachers Partner for Success"

Childlight Conference, Intown Community Schools, "Reading Research-Finding the Pearl in the Oyster", February, 2005.

Using Essential Questions and Enduring Understandings in Curriculum, Chattanooga Christian School, August 7, 2007, all day workshop for the curriculum council.

Professional Development: "Hey, how many points is that book? Motivating Good Readers to Become Great Readers, Brainerd Baptist School, Chattanooga, TN, August 7, 2008

Professional Development: "Spice Up Your Planning Life", Signal Mountain Christian School, August 12, 2008

#### Publications:

Pennington, R. (2005). "Living through" the looking glass: The legacy of Louise Rosenblatt's transactional theory of reader response for Christian language arts Teachers". *Journal of the International Community of Christians in Teacher Education*, 1(1). Retrieved from <http://www.icctejournal.org/ICCTEJournal/past-issues/volume-1-issue-1/living-through-the-looking-glass>

Pennington, R. (2006, December). Children's literature and God's creative nature. *byFaith*, 12. Retrieved from <http://byfaithonline.com/page/arts-culture/childrens-literature-and-gods-creative-natu>

Reviewer for Kiefer, B. (2007). *Charlotte Huck's children's literature*, 9e, New York, NY: McGraw-Hill.