ARE THEY LEARNING? MEASURING PRESERVICE TEACHERS' SKILLS AT DETECTING EFFECTIVE TEACHING INTERACTIONS

A Dissertation

Presented to

The Faculty of the Curry School of Education

University of Virginia

In Partial Fulfillment

of the Requirements for the Degree

Doctor of Philosophy

---

by

Peter D. Wiens, B.A., M.A.

May, 2013

ABSTRACT

Teachers are important for the success of students in schools. Research indicates that teachers are the most important in-school factor in determining student achievement. Traditional teacher education programs prepare the majority of new teachers for service in the K-12 school system. While there is empirical support that teacher education can make a difference in individuals' teaching effectiveness it is not clear what the best way to train teachers might be. There are many similarities among teacher education programs, yet each program can adopt varied approaches and program designs. The implementation of standardized measures that can parse out the effects of teacher education programs on preservice teacher learning could be used to compare the effectiveness of different teacher education models and begin to build knowledge of the effects of those models.

This study examined the efforts of one teacher education program to assess preservice teachers' ability to detect effective teaching interactions in short video clips of preschool language arts classes. The Video Assessment of Interactions and Learning (VAIL) was implemented at three different points in a five-year bachelors plus master's degree program and twice in a two-year postgraduate master's degree teacher education program. The VAIL is a standardized measure based on the Classroom Assessment

Scoring System (CLASS), a standardized observation measure which assesses the quality of teacher-student interactions.

Analysis of three years of data collected at a teacher education program included descriptive analysis, regression analyses, and analysis of variance. Findings indicate that it is possible to measure preservice teachers' ability to detect effective interactions in video recordings and that scores on the VAIL change over time. However, this ability is not predicted by limited individual and programmatic characteristics. Analyses also indicate that the ability to detect effective teaching interactions is associated with observed teaching performance. Taken together these findings provide preliminary support for the VAIL as a standardized measure of teacher education.

Department of Curriculum Instruction and Special Education
Curry School of Education
University of Virginia
Charlottesville, Virginia

APPROVAL OF THE DISSERTATION

This dissertation, Are They Learning? Measuring Preservice Teachers' Skills at Detecting Effective Teaching Interactions, has been approved by the Graduate Faculty of the Curry School of Education in partial fulfillment of the requirements of for the degree of Doctor of Philosophy.

Susan Mintz, PhD., Co-Advisor

Jennifer LoCasale-Crouch, PhD., Co-Advisor

Sandra Cohen, PhD., Committee Member

Jamie DeCoster, PhD., Committee Member

March 20, 2013 Date

DEDICATION

For Annie, Jason and the Chou-Wiens on the way.

ACKNOWLEDGEMENTS

I would like to thank all those people that have contributed to this dissertation and my growth and learning over the course of my doctoral program. First and foremost I have to thank my wife, Annie, who has supported me emotional, intellectually, and financially for the past four years. Without her, none of this would have happened. I can't thank her enough for taking on the role of single-parent and provider while I was away on my studies. Without her continual support and encouragement, especially during the dissertation process, I would not have been able to complete my doctoral program.

My sincerest thanks also go to the Curry faculty that have mentored and taught me during my doctoral program. Specifically, I thank Dr. Susan Mintz, my academic advisor and dissertation co-chair for her academic and professional guidance throughout my years at Curry. Dr. Mintz's support, encouragement, and modeling of excellent teaching have shown me the kind of professor I would like to be someday. I'd also like to thank Dr. Jennifer LoCasale-Crouch, dissertation co-chair and unofficial mentor to me. Dr. LoCasale-Crouch's patience and support while working together with me on projects have been instrumental in my growth as a scholar. I would also like to thank Dr. Sandra Cohen for her guidance and opportunities to work with teacher education data and learn

many valuable things about the field. Finally, thank you to Dr. Jamie DeCoster who is always patient and helpful working with me whenever I ask.

Completing a doctoral program is a group effort. While it may be my name on the diploma, I could not have done this without the support, patience, and love of my family and the faculty of the Curry School.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Teacher education programs train prospective teachers in the necessary skills and knowledge to successfully mediate the classroom environment. The majority of new teachers graduate from these programs (National Research Council, 2010) and yet there is a lack of standardized measures that can provide an assessment of the ability of teacher education programs to prepare individuals for teaching (Zeichner, 2005). A measure that can be implemented reliably across diverse settings, account for individual characteristics, and measure growth in teaching abilities could provide meaningful data for informing teacher education program development in promoting technical skills.

Teachers are important. Within the K-12 public school system in the United States, research has coalesced around the finding that teachers are the most important predictor of student academic success as measured by standardized tests (Goldhaber, 2007; Nye, Konstantopoulos, & Hedges, 2004; Rivkin, Hanushek, & Kain, 2005; Rockoff, 2004). The federal government has recognized the importance of teachers and included a requirement in the landmark No Child Left Behind Act (U.S. Department of Education, 2001) that all teachers should be highly qualified as determined by education degrees and state licensure (US Department of Education, 2001). NCLB legislation

represents the culmination of a shift of accountability for student academic success from families to teachers (Cochran-Smith & Lytle, 2006). Meanwhile teachers today are expected to teach an increasingly diverse group of students in specific content and processes mandated by state boards of education (Bransford, Darling-Hammond, & LePage, 2005).

To improve teachers' abilities to meet the demands of classroom teaching, teacher training can take place at different points in the teaching trajectory. With teachers already in the profession, effective professional development (Ball & Cohen, 1999; Garet, Porter, Desimone, Birman, & Yoon, 2001) and mentoring (Davis, 2010; Ingersoll & Smith, 2004) can improve teachers' instructional abilities. However, effectively preparing novice teachers prior to obtaining teaching positions may provide a basis for increased teaching effectiveness throughout the early years of teaching (Attebery, Loeb, & Wyckoff, 2012), raise achievement levels of students with novice teachers (Attebery et al., 2012), increase teacher confidence and self-efficacy (Darling-Hammond, Chung, & Frelow, 2002), and prolong teaching careers (National Commission on America's Future, 2003). Training of preservice teachers in traditional teacher preparation programs, alternative certification programs, and teacher residencies can help to ensure high quality teachers are present in classrooms by teaching preservice teachers the skills and knowledge required for success in the classroom.

Teacher education has become institutionalized primarily in university settings (Labaree, 2008). There has been a recent growth in alternative certification programs in the United States, but a majority of new teachers still graduate from traditional college and university-based teacher education programs (Zumult & Craig, 2005). While 70-

80% of teachers are trained in traditional teacher education programs, there is variety in this category (National Research Council, 2010). Traditional teacher education programs, housed in universities and colleges, include four year degrees, five year combined bachelor's and master's degrees, and one and two year post graduate degrees (Ludwig, Kirshstein, Sidana, & Bae, 2010). Additionally, within each of these teacher education programs different licensure specialties (i.e. elementary, special education, English, etc.) may have different coursework. For example, the largest United States teacher education accreditation agency is the National Council for the Accreditation of Teacher Education (NCATE) which has separate accreditation program standards for each of twenty-three teaching specialties (National Council for the Accreditation of Teacher Education [NCATE], 2008). Each teaching specialty can be judged separately, leaving the possibility of differences developing between the various teaching disciplines.

Traditional teacher education programs can make a difference in how effectively individuals can impact student learning (Konold et al., 2008). Determining the most effective method for training preservice teachers is an important part of designing and modifying teacher education programs for successfully preparing novice teachers. Analyzing the relative effectiveness of different teacher education programs would improve with standardized measures that can be used reliably across settings. Standardized measures allow the ability to analyze preservice teachers in each of the programs and specialties, compare across programs, and—with prior characteristics—judge how these programs support individual preservice teacher learning across a variety of contexts.

Introduction to the Problem

Measurement using standardized assessments in teacher education settings is important to understand how effectively preservice teachers are prepared to increase student achievement in classrooms. However, there is a lack of standardized measures that can be implemented in teacher education settings (Zeichner, 2005). This section will examine why there is a lack of such assessments. First, it will examine the complexity of teaching that makes measurement difficult in all teaching settings. Then there will be a discussion of research that has led to standardized measures in teaching using direct observation of classrooms. Finally, the difficulties of measuring teacher education will be presented with a look at some strategies that have been used to measure teacher education.

The Complexity of Teaching

Teaching is a complex profession requiring a plethora of skills and knowledge to be done competently (Anderson, Evertson, & Brophy; 1979; Lampert, 1985; Roth, Masciotra, & Boyd, 1999). Teachers that effectively help their students learn have a set of knowledge, skills and dispositions that enables them to promote classroom learning. Effective teachers understand the socio-emotional and cognitive needs of their students (Hamre & Pianta, 2007) while having a deep understanding of pedagogy, subject area knowledge, and the unique pedagogical tools most appropriate for teaching a given subject (Shulman, 1987). Additionally good teachers have a thorough knowledge of the established curriculum and enact that curriculum in ways that make it accessible for their students (Thornton, 2008). They also make solid instructional decisions and implement appropriate models of instruction for the curricular material (Estes, Mintz, & Gunter,

2011). This knowledge coincides with a set of skills that teachers employ to promote, foster, and sustain learning. In so doing, teachers develop skills in relating effectively with their students (Hughes, Cavell, & Willson, 2001). Finally, teachers must develop reflective habits that allow them to continually engage in professional improvement that takes advantage of the contextual factors in which they work (Hartford & MacRuairc, 2008; Schon, 1987).

The Difficulties Inherent in Measuring Teaching

Standardized measures can begin to unpack the complexities of teaching by providing instruments that can differentiate between different levels of performance while being consistent across settings and contexts. A standardized measure needs to be equally reliable in a low-income inner city school as it is in a high-income suburban setting while providing data that allows comparison between the two environments. As John Easton (2012), the director of the Institute of Education Sciences, puts it

> Good measurement helps us identify, define and clarify the nature of a problem. It can help us to determine the linkages among other behaviors and conditions. It can help us communicate across stakeholder communities. Measurement isn't just about outcomes, but also about the processes that we need to improve them (p. 11).

While it may not be possible to measure all of the aspects of effective teaching in one assessment, parsing out important parts of effective teaching can provide a window into teachers' knowledge, skills, and dispositions. A standardized measure based on empirically and theoretically supported components of effective teaching can provide rich data for the evaluation of teachers and the construction of targeted teacher training activities. Some theorists reject the idea of creating standardized measures that quantify phenomena across diverse settings because general data do not apply to individuals and

each context is different (Guba & Lincoln, 1985). Standardized measures do not eliminate the need for in-depth study of the uniqueness of individuals and specific contexts, but it is possible to create standardized measures that can be implemented successfully in different contexts and provide meaningful data for analysis and interpretation.

School and district efforts to evaluate the effectiveness of teachers provide a case study in the difficulties of measurement. For example, the most common form of teacher evaluation is conducted by administrators who visit and evaluate classrooms. Even though state boards of educations set teacher evaluation policy (National Council on Teacher Quality [NCTQ], 2012) this form of evaluation has not been implemented in a standardized way (Jacob & Lefgren, 2005; Peterson, 2004; Weisburg, Sexton, Mulhern, & Keeling, 2009). From a measurement perspective, this means the traditional evaluation model does not provide reliable evidence of teacher performance. In a study of approximately 15,000 teachers evaluated by administrators without a standardized measure, Weisberg and colleagues (2009) found a lack of variation in teacher evaluations whereby 99% of teachers were graded as satisfactory. This finding conflicts with analyses of teacher effectiveness using student achievement test data that have consistently found significant variation between teachers (Ballou, Sanders, & Wright, 2004; Nye, Konstantopoulos & Hedges, 2004). Additionally, the subjective nature of this unstandardized type of measurement makes comparisons between school settings impossible (Peterson, 2004). There is evidence, however, that school districts that devote the resources to it can implement standardized evaluation systems (Odden, 2004). Research on two school districts that implemented a standardized evaluation system

based on the Danielson Framework (Danielson, 1996; Danielson & McGreal, 2000) have shown that this model can be implemented reliably by measuring variation in teacher performance and a correlation with student achievement evaluation models (Broman & Kimball, 2005; Kimball, White, Milanowski, & Borman, 2004; Milanowski, 2004)

Educational researchers rather than evaluators, have had more success in developing standardized measures that can be implemented across settings reliably. One approach to understanding effective teaching has been to focus on the behaviors, or actions, that teachers do in classrooms that can be shown to increase student learning through direct observation of teaching. This idea dates back at least to the 1920s and has taken different names, but the basic idea is to observe what effective teachers do and then teach preservice teachers to implement those same behaviors (Zeichner, 2012). Studying teacher behaviors became prominent under the label of process-product research (Brophy & Good, 1986; Dunkin & Biddle, 1974; Flanders, 1970; Medley & Mitzel, 1963) and has been influential in the science of understanding the actions teachers do in the classroom that enhances student learning. Process-product research has been criticized as being too narrowly focused and missing many of the important aspects of effective teaching (Berliner, 1979; Erickson, 1986). Macmillan & Garrison (1984) criticize the dehumanization of education in the process-product research by saying,

> But teaching is a human activity, and like all human activities it is intentional, a matter of moods and tenses, aspirations, beliefs, and goals. The failure of process-product research to come to grips with essential intentionality of teaching is its greatest conceptual shortcoming. (p. 18).

Additionally, process-product research ignores the importance of the context (Hallinger & Murphy, 1986; Putnam & Borko, 2000) as well as culture (Horowitz et al., 2005; Rogoff, 2003) of the learning environment. In response to such criticisms, Gage and

Needels (1989) wrote, “In focusing on process-product research, we make no claim that it is the best kind of research on teaching or that other kinds of research are not worth doing” (p. 254).

What process-product research has accomplished is to lay the foundation for establishing a scientific basis for understanding what teaching behaviors contribute to student learning (Medley & Crook, 1980). It also helped to build a scientific tradition of standardized classroom observations (Brophy & Good, 1986). Based on correlational analyses, process-product provided an explicit empirical link between what teachers do and what students learn. Teachers that knew and enacted these behaviors could improve student learning. Process-product research also demonstrated that standardized observation measures could be implemented reliably across contexts.

One approach to the direct observation of teaching, The Classroom Assessment Scoring System (CLASS), takes into account a variety of factors within the learning environment. The CLASS acknowledges that the same behavior may have different meaning in different settings and therefore it focuses on the interactions between teachers and students (Pianta & Hamre, 2009b). Instead of focusing solely on the actions of teachers like much of process-product research had done, CLASS examines the behaviors of teachers and students in the context of the classroom environment to assess the quality of the teaching interactions.

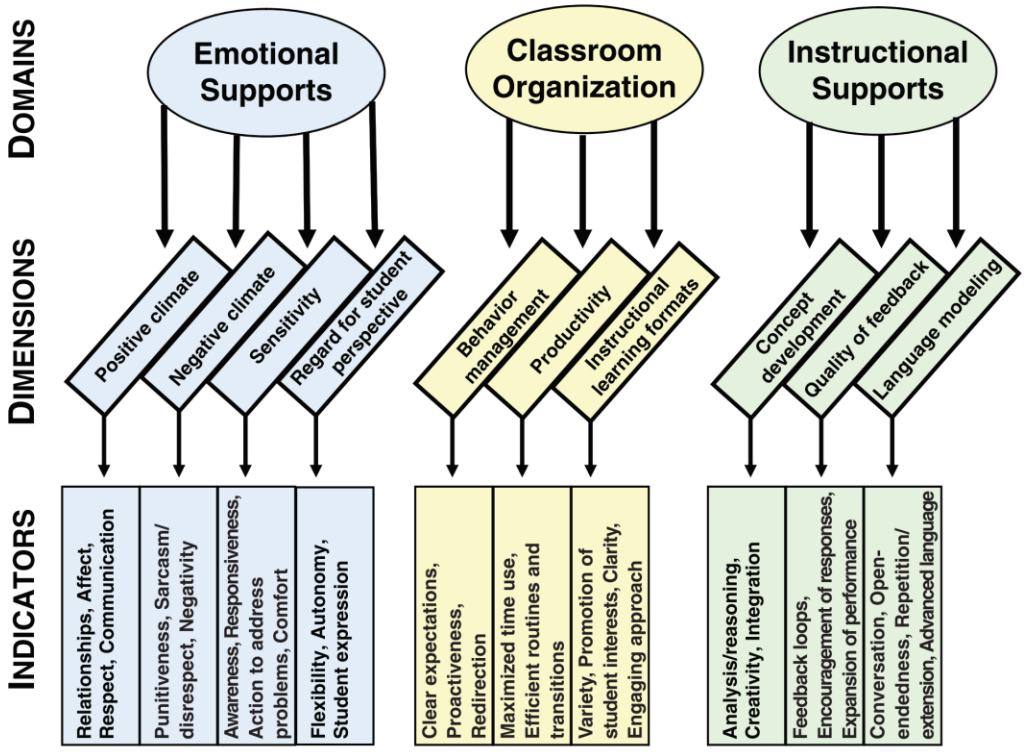
CLASS was developed through extensive observations of classrooms to understand which teacher-student interactions can most impact student learning (Pianta, 2003). Proximal processes-interactions that take place regularly, repeatedly, and over an extended time- serve as the primary source of children’s development (Bronfenbrenner,

1993). Thus, the CLASS focuses on student-teacher interactions as a conceptual framework for understanding and analyzing teacher behaviors that contribute to student learning (Pianta, La Paro, & Hamre, 2008). CLASS organizes teacher-student interactions into three domains that show positive impacts on student learning: emotional support, classroom organization, and instructional support (Hamre, Pianta, Mashburn, & Downer, 2007). As illustrated in Figure 1, within each domain specific behaviors have been identified that contribute to improved student learning. CLASS provides not only a framework for understanding important teaching behaviors, but also a standardized measurement tool for understanding teacher-student interactions that encourage student learning (Hamre et al., 2007). Research has shown CLASS to be a reliable and valid measure that has been used in classrooms at all grades levels both nationally and internationally (Cadima, Leal, & Burchinal, 2010; Graue, Rauscher, & Schefinski, 2009: La Paro et al., 2009; Parkarinen et al., 2010). The Measures of Effective Teaching study, funded by the Bill and Melinda Gates Foundation, adopted CLASS as one of its standardized observation measures and found that CLASS correlated with student achievement gains as measured by achievement tests (Gates Foundation, 2012).

Studying Teacher Education

The complexities of teaching make the creation and implementation of standardized measures difficult. Likewise, creating standardized measurement tools for examining teacher education is also difficult for a variety of reasons. The complexity of teaching makes it impossible to create a valid, reliable, and practical measure that captures all of the knowledge, skills, and dispositions required for effective teaching.

Figure 1
The CLASS Framework for Classroom Interactions (Pianta & Hamre, 2009, p. 111)



Also, there needs to be general agreement as to the outcome of interest. In other words, measures of teacher education effectiveness need to have a clear and agreed upon conceptualization of what constitutes effective teaching. One view of effective teaching comes from Hanushek (2002) who argued that an effective teacher is one who raises students' test scores. Yet another view of an effective teacher is one who uses teaching to change society for the better (Friere, 2009). Yet others may agree with Hamre and Pianta (2007) who lay out an argument that effective teachers are those who meet the socio-emotional and cognitive needs of their students. Even with documents such as the InTASC Model Core Teaching Standards (Council of Chief State School Officers [CCSSO], 2011), teacher education programs may have different foci in what is most emphasized as the core of effective teaching. Additionally, there is the added complexity

of having different teacher education program formats that can impact the ability of standardized measures to be implemented in different settings.

Training preservice teachers to develop the knowledge, skills, and dispositions of effective teaching is equally complex as teaching itself. Teacher education programs are tasked with training novices to be competent teachers in an environment of increasing skepticism towards the abilities of traditional programs to accomplish this lofty goal (Cochran-Smith & Zeichner, 2005; McNergney & Imig 2006). Most, if not all, traditional teacher education programs rely on a similar format of general education coursework, subject area courses, pedagogical courses, and clinical or field experiences (Murray, 2008) to train teacher education students. While subject to state and national accreditation requirements, teacher education programs do have the freedom to create their own conceptual framework (Wilson, Floden, & Ferrini-Mundy, 2002) and between states there is significant variation in course requirements (Constantine et al., 2009). A standardized measure that can be implemented reliably in programs that require different numbers of credit hours and are based on different conceptual frameworks could provide the comparative data necessary to begin to assess the relative quality of requiring more or less credit hours or using different conceptual frameworks.

The field of teacher education research lacks common measures and a common language that help scholars and teacher educators to understand and communicate efficiently (Grossman & McDonald, 2008). Without the standardized measures and consistent language discussed by Grossman and McDonald (2008) researchers and teacher educators have no way to begin to understand how different programs prepare teachers relative to other programs. Without these measures, it is impossible to

empirically identify exemplary programs. To understand the efficacy of different teacher education programs, it is important to develop practical and applicable standardized measures that will illustrate differential effects. The lack of standardized measures that can be used across teacher education programs could be addressed by the collection of rich qualitative data on each teacher education student but this would be time-consuming and impractical. The increased attention on the InTASC standards may begin to develop a common language of effective teaching but, unfortunately, this document does not provide a measure for evaluating teaching or the training of teachers.

Even though there is need for stronger measurement capabilities by implementing standardized measures in teacher education settings, there has not been wide implementation of such measures across different teacher education programs. Current measures of preservice teacher learning that have the potential to be constructed in a standardized way include field experience evaluations, portfolios, and teacher performance assessments. With an increasing emphasis on early field experiences, field experience evaluations completed by university personnel and cooperating teachers might also be used to examine preservice teacher learning (Grossman, 2010). National organizations such as the National Council for the Accreditation of Teacher Education (2008) and the Association of Teacher Educators (2000) have designed and published standards for field experiences that include guidelines for assessment. However, current field experience evaluations are not generally standardized across institutions and are often implemented inconsistently within institutions (Greenberg, Pomerance, & Walsh, 2011). State boards of education set minimum requirements for clinical preparation hours which may cause some standardization of field experiences within states, but there

is great variety between states in these requirements (American Association for Colleges of Teacher Education [AACTE], 2010). The Teacher Work Sample, which was implemented by a consortium of teacher educations schools called the Renaissance Group, provided a standardized way to evaluate student teaching (Henning & Robinson, 2004). However this measure is cumbersome to both preservice teachers and teacher education faculty. In the same vein, the Performance Assessment for California Teachers (PACT) is another standardized measure of student teaching performance (Darling-Hammond, 2006) that will be discussed in detail below.

Portfolios are used in many teacher education programs. Portfolios can be used effectively for documenting preservice teacher learning and readiness to teach (Anderson & DeMeulle, 1998; Wolf & Dietz, 1998; Zeichner & Wray, 2001). The addition of video recordings of teaching into portfolios can show authentic teaching performance and allow for meaningful reflection from preservice teachers (Bannink, 2009). The creation and implementation of portfolios can be standardized as well (Berrill & Addison, 2010; Tigelaar, Dolmans, Wolfhagen, & van der Vleuten, 2004); however, there is no evidence of a single standardized portfolio assessment model being implemented across teacher education programs. The lack of a standardized portfolio model gaining traction across teacher education settings signals that these measures may result from the resource and time intensive nature that implementing a standardized model reliably across settings would involve.

Teacher performance assessments create a standardized measurement tool consisting of multiple measures. The edTPA is an example of such a standardized approach to collecting and examining evidence of preservice teacher learning and

performance. The edTPA developed from the Performance Assessment for California Teachers (PACT; Pecheone & Chung, 2006) which was developed by a consortium of California universities following state legislation mandating teacher education programs implement performance assessments (Darling-Hammond, 2006). The PACT and the edTPA are teaching specialty-specific and consist of two sets of measures: the capstone assessment and embedded signature assessments. The embedded signature assessments are created by each teacher education program and are not standardized across settings. These assessments are designed to be formative and occur throughout the teacher education program. Examples of these assessments include case studies, lesson or unit plans, analysis of student work, and observations of teaching (Pecheone & Chung, 2006). The capstone assessment is a standardized summative assessment of student teaching using an electronic portfolio (Darling-Hammond 2006). Darling-Hammond (2006) describes the capstone assessment components as

1. a description of their teaching context, including students and content;
2. a set of lesson plans from the segment of instruction;
3. one or two videotapes of instruction during the unit (depending on the field);
4. samples of student work during the unit; and
5. written reflections on instruction and student learning during the unit (p. 130).

The collected evidence is then reviewed by trained raters who evaluate and score the preservice teacher's performance. The capstone teaching assessment and the embedded signature assessments are the central framework for the edTPA.

The edTPA is only beginning to be implemented in many states. The edTPA is a promising approach to measurement in teacher education, but based on research on the PACT (from which it was designed) it is expensive (Guaglianone, Payne, Kinsey, & Chiero, 2009), time consuming (Okhremtchouk et al., 2009) and has only been

implemented in certain states. The edTPA does provide comparative data through the capstone assessment allowing teacher education programs to determine which areas preservice teachers are weakest and adjust their programs accordingly (Pecheone & Chung, 2006). However, it does not measure preservice teacher learning. The edTPA provides an authentic measure of a preservice teachers' preparedness to teach, it does not provide evidence of how much a preservice teacher has learned during a teacher education program.

Perhaps the most difficult aspect of creating and implementing standardized measures in teacher education settings is isolating the casual effects of the program from other factors. Preservice teachers bring their own beliefs, abilities, and perspectives with them when they enter a teacher education program (Decker & Rimm-Kaufman, 2008; Dedeoglu & Lamme, 2011; Garmon, 2004). For example, there are clear connections between personal and family characteristics and academic achievement (Clotfelter, Ladd, & Vigdor, 2006; LoGerfo, Nichols, & Chaplin, 2006; Reardon, & Galindo, 2009). An example is a study conducted using national longitudinal datasets, where LoGerfo and colleagues (2006) examined the difference between achievement based on gender and found that boys generally learn more math and girls learn more reading during elementary and high school. Preservice teachers also bring their own ideas about teaching which impact how they respond to teacher education training (Adler, 2008; Lortie, 1975; Grossman, 1990). Research on teacher education can begin to account for these characteristics in a line a empirical inquiry.

To understand teacher learning in a teacher education program, a measure should account for the prior characteristics of the preservice teachers. Examining prior

characteristics can also shed light on how different individuals respond to the construction of a teacher education program; however, the use of standardized measures in teacher education is in its infancy. Therefore, focusing on a few very clear demographic characteristics such as gender, race, and academic aptitude may provide a good starting point for this line of research. Research indicates that gender and race has historical been related to academic achievement (Corbett, Hill, & St. Rose, 2008). Likewise, gender (Duffy, Warren, & Walsh, 2001; Good, Sikes, & Brophy, 1973) and race (Hawkins, Von Cleve, & Catalano, 1991; Zimmerman, Khoury, Vega, Gil & Warheit, 1995) differences also may impact teacher student interactions in the classroom. Controlling for differences in gender, race, and academic aptitude is an important aspect for measuring preservice teacher learning in a teacher education program.

Statement of the Problem

In order to understand and improve teacher education, it is important to measure the effects of teacher education programs on preservice teacher learning. There are many different ways to conceptualize important components of effective teaching. Focusing on teacher-student interactions, however, provides a powerful way to understand effective teaching that has been linked with improved student learning (Pianta, Belsky, Vandergrift, Houts, & Morrison, 2008). In order to understand the impacts of the teacher education program on a preservice teachers' ability to employ effective teacher-student interactions, a distinction must be made between the learning that occurs as a result of the program and not as a result of prior knowledge on the part of the preservice teacher. Because preservice teachers enter teacher education programs with differing skills, knowledge, and background (Decker & Rimm-Kaufman, 2008; Dedeoglu & Lamme, 2011; Garmon,

2004) it is necessary to implement standardized measures in a way that isolates the effects of the training program. One viable approach is to examine the growth in preservice teachers' ability or knowledge by assessing them at the beginning and again at the end of the program in a pretest-posttest design (Creswell, 2008). A standardized measure should also demonstrate sensitivity to differences in individuals that may be a result of these different characteristics upon entrance into a teacher education program. Testing the impact of specific individual characteristics on performance on a standardized measure is a key component to the analysis.

A standardized measure should be able to detect preservice teachers' ability to identify specific teaching interactions that have been shown to produce student learning gains (Hamre et al., 2012). A standardized measure of preservice teachers' abilities to detect effective teaching interactions that is implemented in a systematic manner can produce this evidence of teacher knowledge and learning. When such a measure is administered at multiple time points in a teacher education program it is possible to control for personal abilities upon entrance to the program and isolate the growth in performance. Performance on such a measure, in conjunction with data on preservice teacher demographics and aptitude, could serve to demonstrate how teachers' characteristics potentially impact teacher growth through a teacher education program.

Observing teaching performance of preservice teachers using a standardized measure provides the most direct measure of their teaching abilities. Most teacher education programs, however, lack an authentic teaching experience at the very beginning of the program (American Association of Colleges for Teacher Education [AACTE], 2010). Therefore, standardized observation measures such as CLASS may not

be useful for assessing growth in preservice teachers' ability to implement positive teacher-student interactions because preservice teachers do not have enough opportunities to teach independent lessons in real classroom settings. With a lack of ability to collect evidence of teaching performance early in a teacher education program, alternative measures should be implemented that are linked to valid measures of teaching practice.

Measuring preservice teachers' ability to recognize effective teacher-student interactions may serve as a proxy for skill in teaching behaviors (Jamil, Sabol, Hamre, & Pianta, under review). Teachers that have a demonstrated ability in detecting effective teacher-student interactions have also been shown to exhibit those same behaviors in their teaching practice (Hamre et al, 2012; Jamil, Sabol et al., under review). Understanding the relationship between a measure of teaching ability (CLASS) and the ability to effectively identify effective teaching interactions is a key component to implementing a measure that serves as a proxy for skills in teaching interactions.

The Video Assessment of Interactions and Learning (VAIL) was developed to assess the ability of teachers to detect effective interactions in video-recorded segments of in-service preschool teachers. Based on the CLASS framework, the VAIL focuses on participants' ability to detect effective interactions in three areas that have been shown to matter to student learning (Jamil, Sabol et al., under review) and are elaborated in Table 1: regard for student perspectives, instructional learning formats, and quality of feedback. Taken together, preservice teachers' ability to detect effective interactions can be measured using the VAIL.

Table 1
Descriptions of VAIL Dimensions (Hamre & Pianta, 2007, p. 59)

Domain	Dimension	Description
Emotional Support	Regard for Student Perspectives	The degree to which the teachers' interactions with students and classroom activities place an emphasis on students' interests, motivations, and points of view, rather than being very teacher driven. This may be demonstrated by teachers' flexibility within activities and respect for students' autonomy to participation in and initiate activities.
Classroom Organization	Instructional Learning Formats	The degree to which teachers maximize students' engagement and ability to learn by providing interesting activities, instruction, centers, and materials. Considers the manner in which the teacher facilitates activities so that students have opportunities to experience, perceive, explore, and use materials.
Instructional Support	Quality of Feedback	Considers teachers' feedback focused on expanding learning and understanding (formative evaluation), not correctness or the end product (summative evaluation)

Purpose of the Study

Since teachers are critical to student success in schools (Goldhaber, 2007; Nye et al., 2004; Rivkin et al., 2005; Rockoff, 2004), it is necessary to ensure that teachers are trained effectively. To understand the effectiveness of teacher education programs it is important to know how preservice teachers' abilities change through the program and how these changes are impacted by preservice teacher characteristics. The purpose of this study is to examine preservice teachers' ability to detect effective teaching interactions that have been shown to improve student learning using a standardized video-based measure. Additionally, this study examines how preservice teachers' ability to detect specific teaching interactions predicts their ability to perform effective teaching interactions as measured with a standardized observation tool. This study uses two standardized tools (VAIL and CLASS) to evaluate how well preservice teachers can

identify effective student-teacher interactions and implement effective student teacher interactions during a traditional teacher education program.

Theoretical and Conceptual Framework

This work is theoretically and conceptually grounded in a series of assumptions related to teacher education. These assumptions informed and framed my approach to developing my research questions and design.

Assumption 1: Learning to teach is complex due to the cluster of skills and knowledge required to perform effectively (Kagan, 1992). However, parsing out specific pieces of effective teaching can provide a lens for examining teaching. Teacher behaviors, one such lens, matter to student learning (Brophy and Good, 1986) and can be measured in a standardized way by examining teacher-student interactions (Pianta & Hamre, 2009).

Assumption 2: Teacher education programs can play a significant role in training teachers to more successfully impact student learning (Konold et al., 2008). While there are alternatives to traditional teacher education programs, the majority of new teachers are prepared through traditional programs (National Research Council, 2010) and therefore warrant specific study.

Assumption 3: Learning about teaching begins long before aspiring teachers enter teacher education programs (Lortie, 1975). They come into their teacher education programs with particular characteristics and concerns. Preservice teachers experience changes in their beliefs, knowledge and skills during their teacher education program (Adler, 2008). Content area coursework, pedagogical coursework, and field experiences all play a critical role in preparing individuals to become effective teachers.

Assumption 4: Isolating and measuring specific skills using standardized tools can begin to build a knowledge base that can inform teacher education program construction and policy (Zeichner, 2005).

Taken together these assumptions frame a vision of the ability of standardized measures to identify and assess meaningful components of teaching that can provide data on preservice teacher learning and teacher education program effectiveness.

Research Questions

Given the importance of measuring the effects of teacher education programs on preservice teachers and the empirical and theoretical support behind using teacher-student interactions as a lens for examination (elaborated in Chapter 2), this field warrants additional study. Therefore, this study examines the following research questions:

1. What is the association between preservice teachers' individual characteristics, teaching specialty, and teacher education program and their ability to detect effective teaching interactions?
2. After controlling for initial skill in the ability to detect effective teaching interactions, do preservice teachers' individual characteristics, teaching specialty or teacher education program predict change in their ability to identify effective teaching interactions?
3. What is the association between preservice teachers' ability to detect effective teaching interactions and their observed teacher-student interactions?

Definition of Terms

Classroom Assessment Scoring System (CLASS). CLASS, a standardized observation instrument, is based on decades of research on behavioral markers of

effective teaching (Brophy & Good, 1986; Pianta & Hamre, 2009a) and uses teacher-student interactions as a conceptual framework for understanding effective teaching (Pianta et al., 2008). CLASS organizes classroom interactions into three domains: emotional support, classroom organization, and instructional support.

Effective Teaching. Effective teaching encompasses a collection of skills and knowledge including expertise in both subject matter and teaching. Additionally, effective teachers understand the cognitive, social, and emotional development and needs of their students while creating a safe, nurturing and cognitively demanding learning environment (Hamre & Pianta, 2007).

Prospective Teacher. This is an individual who may want to become a teacher but has not enrolled in a teacher preparation program yet.

Standardized Measure. A standardized measure is one that is the same for all participants that complete it. The measure should demonstrate adequate reliably when tested in different situations and populations allowing for comparisons between all test participants.

Teacher Education. Teacher education refers to those formal programs that are designed to prepare graduates to assume teaching positions in K-12 schools. These programs are generally housed in colleges and universities. Recommendation from these programs is a key requirement for obtaining a teaching license in most states.

Teacher Education Student/Preservice Teacher. These terms are used interchangeable to refer to individuals enrolled in teacher education programs whether or not they have previous teaching experience.

Video Assessment of Interactions and Learning (VAIL). The VAIL is a standardized video-based measure of a participant's ability to recognize effective teaching interactions (Jamil, Sabol et al., under review). Based on CLASS (Pianta et al., 2008), performance on the VAIL has shown an association with effective teaching interactions with in-service teachers (Hamre et al., 2012; Jamil, Sabol et al., under review) and has demonstrated reliable implementation in a teacher education setting (Wiens, Hessberg, LoCasale-Crouch, & DeCoster, 2013).

Conclusion

This study contributes to a gap in the literature by studying the ability of preservice teachers to detect effective teaching interactions through the implementation of a standardized measure. Using two standardized measures (VAIL and CLASS) as well as demographic and program data this study examines the change in abilities of preservice teachers over time while also assessing the relationship of their abilities to detect effective interactions and implement these interactions in a student teaching setting. The findings from this study inform teacher education programs on the usefulness of a standardized measure to assess the effects of a teacher education program.

CHAPTER TWO

REVIEW OF THE LITERATURE

Teacher education programs are designed to assist future teachers in developing the necessary knowledge, skills and dispositions for effectively teaching students. This study examines preservice teachers' abilities to detect effective teaching interactions in videos of in-service teachers using a standardized, video-based measure. To understand the need for such a study it is first crucial to situate measurement of preservice teacher abilities in the context of broader topics. This chapter explores why it is relevant to study teacher education in the United States, the need for increased measurement capabilities around teacher education, the challenges with measuring effective teaching, the added challenges of measuring teacher education, and the efforts of schools and researchers to address these issues. The literature on these topics provides a rationale for exploring the use of a standardized measurement of preservice teacher competencies. This review of literature is organized into four sections. The first section examines the theory and research related to why teacher education is important to student learning. The second section discusses the literature on the intricacies of skill development in teaching that make it difficult to measure effective teaching. The third section analyzes attempts of schools and researchers to measure effective teaching. The final section expands the

examination of the difficulties in measurement to teacher education students and programs.

The Importance of Teacher Education

The impetus for studying teacher education derives from the important role teacher education plays in the preparation of novice teachers for classroom duty. This section describes research indicating the importance of teachers to student learning and why traditional teacher education is critical to ensuring effective teachers fill K-12 classrooms.

The Importance of Teachers

As early as the 1960s with the famous "Coleman Report" (Coleman, 1966), educational researchers have shown that teachers are the biggest school-based factor in determining student learning. Research also indicates that student learning is an important predictor of life outcomes for students including adopting more progressive views (Kingston, Hubbard, Lapp, Schroeder, & Wilson, 2003) and predicting economic success (Mulligan, 1999; Murnane, Willett, Duhaldeborde, & Taylor, 2000). Drawing on data collected during the Tennessee Star randomized classroom size experiment, (Chetty et al., 2010) found that students in higher quality K-3 classrooms earned more money as adults, were more likely to attend college, save more for retirement, and live in better neighborhoods. The impacts of educational quality can have consequences for students' lives long after the schooling years.

More recent research has strengthened the connection between teachers and student learning (Goldhaber, 2007; Nye et al., 2004; Rivkin et al., 2005; Rockoff, 2004). In fact, Hanushek (1992) finds that the quality of a teacher can make the difference of a

full year's worth of learning for students. The measured effects of the quality of the teacher are far greater than any other school-based intervention (Goldhaber, 2008). Clearly teachers are important to student learning.

Policy makers have also placed an emphasis on teachers with increased attention to creating regulations that focus on teaching effectiveness (National Research Council, 2010). The landmark No Child Left Behind Act (NCLB) mandates all teachers should be highly qualified as determined by education degrees and state licensure (US Department of Education, 2001). While NCLB does set minimum standards for teachers to meet prior to taking over classroom duties, this mandate does not ensure that teachers who meet those requirements are helping students learn. Based on the importance of teachers to student learning, ensuring quality teachers—not highly qualified teachers—in every classroom is an important issue. A quality teacher has internalized an elaborate system of skills and knowledge that is flexible and adaptable in a variety of contexts.

Improving Teaching Effectiveness

Efforts to ensure that teachers are effective can take different forms of training and occur both throughout and prior to the teaching career. Upon hiring of novice teachers, many school systems provide induction programs that have been shown to be beneficial to improving teacher effectiveness. In-service professional development, when structured properly, including coherent content, sustained active learning, and the nurturing of collegial support networks can improve teaching performance (Ball & Cohen, 1999; Garet et al., 2001; Gersten, Dimino, Jayanthi, Kim, & Santoro, 2010; Penuel, Fishman, Yamaguchi, & Gallagher, 2007).

While professional development and induction programs (Davis, 2010; Ingersoll & Smith, 2004) are certainly important, recent research points to the importance of teacher training prior to having classroom responsibilities. Several studies using student achievement data have shown that new teachers experience improvements in effectiveness over the first three to five years of classroom teaching (Boyd, Lankford, Loeb, Rockoff, & Wyckoff, 2008; Clotfelter, Ladd, & Vigdor, 2007; Hanushek, Kain, O'Brien, & Rivkin, 2005; Plecki, Elfers, & Nakamura, 2012). A recent study by Atteberry, Loeb, and Wyckoff (2012) investigated if all teachers' show similar growth over these early career years. They examined ten years of data on over 10,000 teachers in the New York City Public Schools. Through estimating the growth trajectories of the teachers using student achievement data, Attebery and colleagues found that teachers who were more effective when they entered the teaching profession also showed steeper growth trajectories in their teaching effectiveness over the first five years of their career. These results demonstrate the importance of preparing teachers prior to their assuming teaching positions.

There are competing models for training preservice teachers for classroom duty including traditional, university-based teacher education programs, alternative preparation programs, and teacher residency programs. All of these programs aim to prepare teachers that can effectively teach students. Research does not indicate which pathway to teaching is most effective at preparing preservice teachers for the demands of the classroom (Constantine et al., 2009). Studies have shown that student achievement in classes taught by alternatively certified teachers lags in the teacher's first year, but the difference disappears by the second or third year in the elementary grades (Boyd et al.,

2005; Darling-Hammond, Holtzman, Gatlin, & Heilig, 2005; Kane, Rockoff, & Staiger, 2006). However, in a survey of nearly 3000 novice teachers, Darling-Hammond, Chung, and Frelow (2002) found that traditionally certified teachers feel like they were better prepared and had a higher sense of self-efficacy. Work by Kee (2012) confirmed Darling-Hammond et al.'s results; however, a recent study of over 1,100 teachers found that traditionally certified and alternatively certified teachers did not exhibit a difference in self-efficacy (Mueller, 2012) further clouding the debate over which pathway is more effective at preparing teachers.

In spite of growth in alternative certification programs (Humphrey & Wechsler, 2007; Zumult & Craig, 2005), currently 70-80 percent of newly certified teachers have completed traditional teacher education programs including 4-year bachelors programs and 1-year post baccalaureate programs (National Research Council, 2010). This continues a tradition of teacher education being primarily housed in institutions of higher education (Labaree, 2008). These traditional teacher education programs provide a similar framework containing pedagogical coursework, subject area coursework, and clinical or field experiences (Murray, 2008). Boyd, Goldhaber, Lankford, and Wyckoff (2007) report on the requirements imposed by states for teacher education programs and show that most states require secondary preservice teachers to major in their content area, as well as minimum requirements in pedagogy courses and field experiences. Levine (2006) reported that elementary teachers have similar state requirements. All fifty states required specific education course work and field experiences while 36 states prescribe coursework by subject area as well (Levine, 2006).

The importance of teachers to student learning necessitates specific training so that trained teachers can increase student success in academic learning. The importance of teaching effectiveness when a teacher begins his or her career, combined with the fact that a majority of new teachers are trained in traditional teacher education programs makes studying these teacher education programs particularly important. Designing effective ways to measure the quality of teacher education programs may contribute to producing novice teachers that are better prepared to meet the needs of their students and promote learning. In addition, what is learned about the involved process of professional preparation for novice teachers may be applicable to other teacher education settings.

Complexity of Teaching

Prior to designing and implementing standardized measures of teacher education, it is important to determine what effective teaching looks like. Research on teacher education developed out of research on teaching (Grossman & McDonald, 2008) and it is important to base measurement of teacher education on empirical evidence of effective teaching practice. In order to understand the goals of teacher education, it is important to have an understanding of what effective teaching is and how to measure it. Research on teaching has developed a picture of effective teaching that encompasses an array of skills and knowledge (Ball & Cohen, 1999; Brophy, 1999; CCSSO, 2011). The complexity of this set of skills and knowledge (Anderson et al., 1979; Lampert, 1985; Roth et al., 1999), which will be discussed in this section, complicate the task of measurement in the field of effective teaching.

Currently, the dominant discourse in education has been framed around what knowledge, skills and dispositions teachers need to have to be effective (Feiman-Nemser,

2008). Research indicates that one skill set that effective teachers have is the knowledge and skills to meet the socio-emotional needs of all students (Pianta & Hamre, 2009a; Hughes et al., 2001). An examination of interview and academic data from 1,480 middle school students showed that students who felt safe in the learning environment and whether they felt able to meet the challenges presented to them affected their academic success (Roeser et al., 2000). Roeser, and colleagues (2000) point out that,

> The challenge for middle school teachers is not a question of getting students motivated or not but rather, in part, getting them to be motivated to learn rather than motivated to protect themselves from situations they perceive as threatening to their self, meaningless, or somehow threatening to their social image (p. 454).

The need for positive emotional support also extends to feelings of student autonomy. Stefanou, Perencevich, DiCinto, & Turner (2004) define autonomy as actions that are chosen or for which one is responsible (Deci & Ryan, 1987). Assor, Kaplan, and Roth (2002) studied more than 800 elementary and middle school-aged Israeli students and found that students were much more engaged in school in the absence of autonomy-suppressing behaviors.

Effective teachers develop the knowledge and skills to meet the cognitive needs of their students by connecting new information to the student's prior knowledge (Bransford, Darling-Hammond, & LePage, 2005; National Research Council, 2000). In a series of three experiments on undergraduate and graduate university students, Schwartz and Bransford (1998) demonstrated that the degree participants were able to build prior knowledge previous to completing a task made a significant difference in performance on that task. Participants in the study that were aided in building a depth of prior knowledge before hearing a lecture or reading a text scored higher on an assessment one-week later. To meet the cognitive needs of students, teachers must have a deep knowledge of their

subject matter, pedagogical methods, and how to best match subject matter and pedagogical methods for their field (Shulman, 1986). Shulman's concept of pedagogical content knowledge has been adopted by researchers in multiple teaching specialties (Loughram, Mulhall, & Berry, 2004; Masters, de Kramer, O'Dwyer, Dash, & Russell, 2010; Silva & Mason, 2003). Hill, Rowan, and Ball, (2005) conducted a study on the effects of the mathematics pedagogical content knowledge of elementary teachers on their students' achievement. They drew from a national sample of 115 elementary schools from 15 states. After controlling for key student and teacher covariates, the researchers found that teachers who scored higher on a measure of mathematical pedagogical content knowledge had students who scored higher on standardized tests in mathematics (Hill et al., 2005).

Additionally, teachers must understand the established curriculum and enact the curriculum in the manner they believe to be the most effective (Thornton, 2008). Even in settings where researchers are attempting to tightly control curriculum implementation, teachers take on curriculum reforms at varying levels (O'Donnell, 2008). Baker, Kupersmidt, Voegler-Lee, Arnold, and Willoughby (2010) studied the uptake of a curricular implementation with 49 pre-school teachers and found that teachers enacted 70% of the curricular intervention. This was after teachers were paid for summer training. Similar results were reported by Hamre et al. (2010) where 80% of a 30-week pre-school curricular intervention was implemented by teachers even after the teachers had agreed to fully participate. With the growth of standardized curriculums, teachers have had to wrestle with adapting the curriculum to their classrooms (Jones et al., 1999; Stone & Lane, 2003). In a qualitative metasynthesis of forty-nine studies, Au (2007) found that

teachers changed their instructional methods and curriculum goals in order to accommodate new curriculum standards. Therefore teachers play a vital role in determining what content is taught in classrooms which can impact student learning.

Meanwhile, teachers must continually reflect on their own teaching and strive to make improvements in their teaching performance (Schon, 1983, 1987). Reflective teaching as a process of examining teaching processes critically and adjusting future teacher performance can have a beneficial impact on teacher effectiveness (LoCasale, 2005; Loughran, 2002). In a qualitative study of four teachers on the use of reflective journals, the teachers found the reflective experience to be supportive of their growth and helped them to develop a supportive community (Swain, 1998). Baird, Fensham, Gunstone, and White (1991) conducted a three-year naturalistic case study with 13 preservice science teachers, 14 in-service teachers, and 64 students from grades eight through eleven. Baird and colleagues found that reflection led to significant gains in levels of satisfaction and performance of both teachers and students.

Each of the individual skill sets described above need to be part of an elaborate system of effective teaching where teachers become adaptive and flexible experts. However, the research presented here also demonstrates that individual pieces of the more elaborate skill set can be isolated for research and training purposes. The research indicates that teachers can be trained on each of these skills to improve instructional quality and student learning.

The complexity of teaching develops from the constellation of skills and knowledge necessary for teachers to be effective. Effective teachers help their students attend to their students' socio-emotional needs while helping them develop new cognitive

skills. Meanwhile they handle curriculum decisions and continually reflect and adjust their teaching practice. Understanding the different components of effective teaching brings the goals of teacher education into focus—training preservice teachers to have the knowledge and skills to help individual students develop success in their classrooms. This section has shown that much is known about effective teaching and that we can use what is known to develop programs that can help ensure classroom success.

Measuring Teaching

Much as teacher education builds on research in effective teaching, it also builds on research in measuring teaching. The field of teacher education can take lessons from measurement in teaching to understand how to implement and adapt measures for use with preservice teachers. The following discussion examines more closely the need for standardized measures, some examples of current measurement of teaching, and looks specifically at one line of research on standardized observation of teaching.

Multiple standardized measures that can be administered reliably across different contexts and have demonstrated evidence of validity could help to promote a science of teaching. Creation and implementation of standardized measures can strengthen our ability to understand teaching quality, adjust teacher training, and inform education policy (Pianta & Hamre 2009a). Standardized measurement in education is important because it allows communication across settings, defines problems, and clarifies solutions (Easton, 2012). It would allow a variety of teacher education endeavors to have a common vocabulary about which they can discuss the opportunities and challenges in developing new teachers. To achieve this goal and because of the complexity of teaching, it will require multiple standardized measures with different measures capturing different

aspects of teaching (Douglas, 2009). To illustrate this point, two examples of teaching measures will be briefly discussed and will be followed by an examination of one line of research that has informed a standardized measure of teaching effectiveness.

Teacher evaluation in schools provides the first example of the difficulty of measuring teaching quality and provides valuable lessons for measurement in teacher education. Traditionally, teacher evaluation in schools has been performed by school-based administrators who assess teaching performance based on criteria developed by the district's central office that includes subject criteria and some teacher behaviors assumed to be associated with student achievement (Peterson, 2004). Because of the individuality of the thousands of school districts in this country, specific evaluation instruments lack consistency and reliability (Jacob & Lefgren, 2005). The end result of this lack of standardization is that teacher evaluations performed in this manner provide little valid data for either school districts or for teachers (Weisberg et al., 2009). In their study of approximately 15,000 teachers evaluated by administrators, Weisberg and colleagues (2009) found that 99% of teachers were assessed as satisfactory. A system that cannot differentiate between effective and non-effective teachers is not useful for assessing teachers or for constructing appropriate training activities.

Another example of efforts to measure effective teaching is the use of student achievement test score data to estimate the effects of a particular teacher on student learning (Braun, 2005). Value-added models (VAMs) have been controversial (Baker et al., 2010; Glazerman et al., 2010) partly because they rely on longitudinal datasets of student test performance and demographic data to calculate student test score gains compared to predicted test performance. The estimated difference is then considered the

teacher's effect on the student's learning (Sanders, 2000). Because this form of evaluation is based strictly on student learning as measured by standardized tests and it has the ability to factor in student demographic and previous academic performance it has been adopted by a variety of educational researchers (see Boyd et al., 2008; Kupermintz, 2003; Sanders & Horn, 1998, for examples) and increasingly by state boards of education (NCTQ, 2012). The lure of VAMs lays in the quantifiable nature of the resulting estimate and these models are beginning to be turned on teacher education—a topic which will be discussed later in this chapter.

VAMs provide an estimation of the effects of teachers, but they cannot be considered standardized measures of teaching effectiveness because they do not provide any information about what is happening in classrooms. They do not serve to "determine the linkages among other behaviors and conditions" (Easton, 2012, p. 11). Nor do VAMs provide useful information for improving teaching performance or tailoring teacher training to an individual's needs (Papay, 2012). While VAMs may have their purpose, they cannot replace standardized instruments that can measure the teaching procedures that contribute to student learning. They do not provide the information necessary to improve teacher education programs.

Standardized measures are important to education because they provide a way to reliably assess teaching effectiveness across settings. While these measures have existed for comparing students across settings for some time, there is a lack of equivalent measures of teaching effectiveness. Administrative evaluations of teachers and VAMs both provide lessons for developing a standardized measure of teacher education programs—implementation of these evaluation formats in teacher education settings will

be described later in this chapter. There remains a need for additional research on alternative instruments to assess both teacher effectiveness and the programs that educate teachers.

Standardized Observations of Teacher Behaviors

Another approach to understanding teacher effectiveness focuses on direct observation of what happens within classrooms. This line of research dates back nearly a century and has adopted different names over that time, but the foundational idea is to observe the behaviors of teachers that positively impact student learning and then train other teachers to enact those behaviors (Zeichner, 2012). This approach to research became influential under the label of process-product research that has served to empirically build an understanding of what types of behaviors can promote student learning (Brophy & Good, 1986; Dunkin & Biddle, 1974; Flanders, 1970; Medley & Mitzel, 1963).

Process-product research typifies the challenges of standardizing measurement of a complex act. Process-product typically focuses on smaller units of teacher behavior and does not address all aspects of effective teaching which has been criticized by many. A key problem with process-product's narrow focus on teaching behaviors is that it misses much of the complexity necessary for effective teaching (Berliner, 1979; Erickson, 1986; Macmillan & Garrison, 1984). Teaching requires a combination of skills and knowledge that are contextually influenced. And research indicates that context (Hallinger & Murphy, 1986; Putnam & Borko, 2000) and culture (Horowitz et al., 2005; Rogoff, 2003) are important factors in student learning. As Zeichner (2012) summarizes,

> "…there is a danger of narrowing the role of teachers to that of technicians who are able to implement a particular set of teaching strategies, but who do not

> develop the broad professional vision (deep knowledge of their students and of the cultural contexts in which their work is situated), and the relational skills they need to be successful in the complex institutional settings in which they will work (Butin, 2005)" (p. 379).

Process-product researchers are careful to argue that teaching is complex and it is important to conduct many different types of research and process-product is simply one approach that has shown to be useful (Gage & Needels, 1989). Using standardized observation protocols such as the ones developed by the process-product tradition form one way to examine and measure teaching effectiveness. While process-product research does not provide a method for measuring the totality of effective teaching, it has provided a one way to conceptualize effective teaching as a series of high-leverage practices. There is a danger, however, that the demand of classroom life can keep researchers and teacher education curriculum designers focused on these isolated technical skills.

Another approach, that takes into account a variety of factors to direct observation of teaching behaviors, is the Classroom Assessment Scoring System (Pianta, La Paro, & Hamre, 2008; Pianta, Hamre, & Mintz, 2012). CLASS is a standardized observation measure that uses direct observation of teacher-student interactions as a framework for conceptualizing and measuring effective teaching which can be used reliably across grades, teaching specialties, and settings (Pianta & Hamre, 2009a). Similar to process-product research, CLASS was developed out of extensive observations of classrooms (Pianta, 2003) which demonstrated the critical importance of teacher-student interactions to student classroom success. CLASS uses these teacher-student interactions as its conceptual framework and organizes these interactions into three domains: emotional support, classroom organization, and instructional support (Pianta, La Paro, & Hamre, 2008). Each domain is theoretically and empirically supported and divided into

dimensions that are described by specific teaching interactions (Hamre & Pianta, 2007; Hamre et al., 2007).

There are different versions of the CLASS protocol designed for use with different ages of students (toddler, Pre-K, K-3, upper elementary, and secondary). The domains are consistent across all versions of the CLASS; however, there are slight variations in the dimensions (teachstone.org). The Pre-K version is the most relevant to this paper and the following discussion will focus on that version.

Emotional support. The domain of emotional support provides examples of interactions that can promote student learning such as teachers displaying respect for students, supporting student autonomy, and creating emotionally safe learning environments. Emotional support, as conceived by CLASS (Pianta, La Paro, & Hamre, 2008) include classroom climate (positive and negative), teacher sensitivity, and regard for student perspectives. Hamre and Pianta (2007) describe the emotional support domain as grounded in two areas of developmental theory: attachment theory and self-determination theory. Attachment theory (Ainsworth, Blehar, Water, & Wall, 1978; Bretherton, 1985) suggests that as people feel more comfortable and safe in the presence of an attachment figure, they are more willing to explore and take risks. Meanwhile, self-determination theory (Ryan & Deci, 2000) posits that humans need to feel competent, relatedness, and autonomy to function at their optimal level.

There is strong empirical support for focusing on emotional support in academic settings. Students who feel a positive relationship with their teachers tend to have more engagement in school (Decker, Dona, & Christenson, 2007), more satisfaction with school (Baker, 1999), and higher academic achievement (Klem & Connell, 2004). In a

longitudinal study of 880 children from birth through third grade, O'Connor and McCartney (2007) found that high quality teacher-student relationships resulted in higher student achievement and higher levels of student engagement. Students are also more likely to be academically motivated when they feel competent, positively related to others and autonomous (Roeser et al., 2000). Responsive and sensitive caregivers have also been shown to somewhat mediate the effects of exposure to social risks outside of school (Burchinal, Roberts, Zeisel, Hennon, & Hooper, 2006).

Classroom Organization. Classroom organization provides a framework for improving student learning through interactions such as presenting clear behavioral expectations, managing instructional time efficiently, and providing interesting learning activities. CLASS conceptualizes the classroom organization domain as containing the dimensions of behavior management, productivity, and instructional learning formats. "Classroom organization is a broad domain of classroom processes related to the organization and management of students' behavior, time, and attention in the classroom" (Hamre & Pianta, 2007, p. 62). This domain is based in the work of developmental psychologists studying children's self-regulatory skills (Raver, 2004) and how those skills relate to academic success (Hamre, Downer, Jamil, & Pianta, in press).

Empirical support for attention to classroom organization demonstrates its importance to teaching (Emmer & Stough, 2001). Effectively structuring class time to maximize learning opportunities is related to increased student achievement (Brophy & Good, 1986). Teacher classroom management techniques have a causal effect on student behaviors (Arnold, McWilliams, & Arnold, 1998). In a longitudinal study of 275

preschool students Dominguez, Vitiello, Maier, and Greenfield (2010) demonstrated that effectively organized classrooms can also influence student behaviors.

Instructional support. Instructional support provide teachers a template for interactions that support student learning such as promoting higher order thinking skills, using feedback that expands student thinking, and focusing on formative evaluation. The CLASS instructional support domain includes the dimensions of concept development, quality of feedback, and language modeling. Emphasis on instructional support is based on the theoretical foundation of research on children's cognitive and language development (Pianta & Hamre, 2009a). The ability of teachers to structure instruction and help connect content to student's prior knowledge is important to students' mastery of new material (National Research Council, 2000).

Research indicates that teachers who engage their students in higher order thinking activities promote learning (Taylor, Pearson, Peterson, & Rodriguez, 2003; Zohar & Dori, 2003). Specifically, Taylor and colleagues (2003) studied 88 high-poverty classrooms and discovered that teachers that engaged their students in higher order thinking had students that learned more reading skills through an academic year. Assisting students with appropriate scaffolding has also been shown to increase learning (Wharton-McDonald, Pressley, & Hampton, 1998) as well as helping students develop metacognitive skills (Veenman, Wilhelm, & Beishuizen, 2004; Zohar, 1999)

CLASS provides a conceptual framework and a language for thinking about effective teaching that examines the interactions between teachers and students. It is also a standardized measure of teaching effectiveness that has been shown to be related to student learning as measured by standardized tests (Gates Foundation, 2012; Grossman et

al., 2010). While it is only one way to conceptualize effective teaching, it does present an empirically validated assessment that measures high leverage practices which have been used reliably in classrooms across the country (Gates Foundation, 2012; LoCasale-Crouch et al., 2007).

Examining measurement in teaching provides evidence for considering standardized measurement in teacher education. It makes sense to use what we know about effective teachers to design teacher education programs. Though the complexities of teaching make measurement difficult, the lessons to be learned from research on teaching for teacher education are many. They include the knowledge that there is a lack of standardization in teacher evaluation making comparisons difficult across settings and that value-added models provide a measure of teaching effectiveness, but not of actual teaching performance. Meanwhile, standardized observation measures provide a reliable and valid way to measure teaching effectiveness across different grade levels, content areas, and contexts.

Measuring Teacher Education

The need for standardized measurement in teacher education equals that of the need for standardized measures in preschool through high school (PK-12) education. Teacher education programs strive to prepare graduates to effectively manage the complexities of the profession to help their students learn (Darling-Hammond, 2006). The complexities involved in training teachers equal the complexities in the act of teaching itself. Teacher education programs are tasked with training prospective teachers in a society that increasingly doubts their worth in this process (Cochran-Smith & Zeichner, 2005; McNergney & Imig, 2006). In addition to informing policy and

improving teacher education practice, and by extension teaching performance, good measurement within teacher education can provide validation of the role of teacher education to the greater educational endeavor. The purpose of implementing standardized measures in teacher education is three-fold. First, it is critical to have a common vocabulary to discuss the impact of a variety of teacher education approaches. Second, it is important to be able to measure the impacts of teacher education programs on students using reliable measures so that we can know what is happening in our teacher education programs. And, finally, data will be generated between schools and across state and national boundaries that can provide the basis for important programmatic decisions. The use and potential for measures in teacher education must be judged in terms of these three opportunities.

The importance of standardized measures in teacher education develops from program differences and the differences among individual preservice teachers. Without standardized measures it is impossible to compare the progress of one teacher versus another and without standardized measures that can be used across programs, it is impossible to compare one teacher education program to another. As mentioned, traditional teacher education programs account for 70-80% of new teachers (Levine, 2006; National Research Council, 2010) but there is considerable variety within these programs (Zeichner & Conklin, 2008). While based around a similar model of general and subject area coursework, pedagogical course work, and field experiences (Murray, 2008), teacher education programs have the freedom to create their own conceptual frameworks (Wilson et al., 2002).

Within the constraints of national and state accreditation agencies, teacher education programs construct different programs such as four and five year degrees and postgraduate degrees (Goodland, 1990; Levine, 2006; Ludwig et al., 2010). In a survey of 709 members of the American Association of Colleges of Teacher Education Teacher Education (AACTE), Ludwig and colleagues (2010) found that of the 709 institutions that responded, 676 offered initial bachelor's degrees and 540 offered post-baccalaureate initial education degrees. Teacher education programs are regulated largely at the state level. However, states have imposed differing requirements, as Zeichner and Conklin (2008) explain

> State policy contexts also vary in significant ways that help define the character and quality of teacher education programs. For example, a four-year program in Texas where the state has capped the number of education courses in preservice programs will be very different in significant ways from a four-year program in a state like Wisconsin where the state has maintained a strong emphasis on the professional component of preservice programs (p. 270).

In an analysis of different pathways to teaching across the United States, Constantine and colleagues (2009) collected data from teachers in 20 districts located across seven states. The research team found that teacher education programs required significantly different amounts of coursework ranging from 240 to 1,380 hours.

Within each traditional teacher education program there are different programmatic and degree offerings (Ludwig et al., 2010). Additionally, within each of these degree programs there may be different teaching specialties such as elementary education, health and physical education, foreign language, English/language arts, and more. Each of these teaching specialties requires some unique teaching pedagogical and disciplinary content (Krauss et al., 2008; Loughran et al., 2004; Shulman, 1986) that may require differences in program construction at the teaching specialty level. For example,

a study of nearly 700 elementary teachers in 15 states indicated that teachers' mathematical knowledge for teaching was predictive of student achievement in math (Hill et al., 2005). It is important, based on this information, to examine how teacher education programs help novice teachers develop mathematical knowledge. Of course, it can be assumed that there are a variety of other sets of knowledge and skills that are important to effective classroom teaching. The only way to understand how different teacher education programs prepare preservice teachers to effectively teach is through the development and implementation of standardized measures that can be used across disparate settings.

Within teacher education programs, individual preservice teachers also enter with differing personal experiences and characteristics. Preservice students enter teacher education with beliefs, abilities, and perspectives (Decker & Rimm-Kaufman, 2008; Dedeoglu & Lamme, 2011; Garmon, 2004) that have developed over their years of education experience as PK-12 students (Lortie, 1975). Dedeoglu and Lamme (2011) surveyed preservice teachers and found that demographic variables were significantly related to participants' beliefs about diversity. The researchers found that Protestants and Catholics were more conservative and that the amount of intercultural friendships and ethnicity were associated with views on diversity. Preservice teachers also develop ideas about teaching prior to entering teacher education programs that may affect how they internalize the training they are receiving (Adler, 2008; Grossman 1990). A case study of English Education students in an English methods class demonstrated how the preservice teachers had many formulated ideas about the content of English as well as the methods of instruction (Grossman, 1990). Therefore standardized measures in teacher education

must be implemented in a way that is sensitive to the preservice teacher characteristics and beliefs including academic content and diversity prior to beginning the program.

Research on teaching discussed previously demonstrates the social and affective component of teaching is critically important to student learning (Bronfenbrenner, 1993). The classroom context (Hallinger & Murphy, 1986; Putnam & Borko, 2000) as well as teacher and student culture (Horowitz et al., 2005; Rogoff, 2003) also play a role in student learning. Therefore, it may be helpful to focus on these relational aspects of preservice teachers to understand their impact on preservice teacher learning and performance. However, with current measures of teacher education in their infancy, it is more important at this point in the empirical process to focus on demographic and aptitude of preservice teachers. This allows for a more parsimonious model that allows examination of the sensitivity of the measure to these specific individual characteristics.

There are differences in academic performance between different gender and racial groups (Corbett et al., 2008). Research also indicates that demographic characteristics can impact teaching practice. Gender is one demographic characteristic that has been examined by researchers and been shown to affect teaching practice (Duffy, Warren, & Walsh, 2001; Good, Sikes, & Brophy, 1973). Bennett (1982) examined student evaluations of college professors and found that female instructors were perceived as warmer and more potent individuals. Ethnicity and racial experiences can also impact teachers' views on educational issues (Dedeoglu & Lamme, 2011; Garmon, 2004). Likewise, there is evidence that racial and ethnic differences between teacher and students result in differing ideas about school behavior (Hawkins, Von Cleve, & Catalano,

1991) and increased disciplinary rates for minority students (Zimmerman, Khoury, Vega, Gil, & Warheit, 1995).

In addition, academic ability as measured by standardized exams (e.g. SAT, GRE, and licensure exams) has been shown to predict teacher effectiveness using student achievement data (Clotfelter et al, 2006; Ferguson & Ladd, 1996; Goldhaber, 2007). In a study of preservice teachers in a five-year bachelor's plus master's degree at the University of New Hampshire, Andrew, Colbb and Giampietro (2005) found a relationship between GRE scores and performance in a one-year field placement. This research suggests that demographic features such as gender and past academic achievement of preservice teachers must be accounted for in measuring teacher education.

Measures that do not in some way control for this variation across individuals cannot provide complete data on the impacts of the teacher education program because these measures can be confounded by individual characteristics. One approach to this problem is using pretest-posttest measures (Creswell, 2008). By administering an assessment at the beginning of a problem, researchers can capture the knowledge and abilities of entering preservice teachers and then control for this ability in the end of course posttest. Additionally, the measure needs to be sensitive to changes in preservice teachers' skills and knowledge over the course of the teacher education program. The following section will describe some examples of measurement in teacher education and discuss the strengths and weaknesses of each.

Measures of Teacher Education

Understanding how different program constructions and frameworks interact with a variety of preservice teacher beliefs, knowledge, and skills could provide useful

information for teacher education programs wishing to improve their programs and to policy makers interested in setting policies that promote the training of effective teachers. The field of teacher education does have some experience with the creation and implementation of standardized measures. Four current options for measurement in teacher education are portfolios, field experience evaluations, teacher performance assessments, and use of student achievement data.

Teacher portfolios are a collection of artifacts that are selected by the preservice teacher to demonstrate learning and competence in teaching.

> In preservice teacher education programs the teaching portfolio offers opportunities for student teachers' experiences, thoughts, actions, and subsequent learning about teaching to be documented. Therefore, these shaping forces which impact on a student teacher's learning can also serve to help them better articulate their own developing professional knowledge (Loughran & Corrigan, 1995, p. 565).

Teacher education programs use portfolios to encourage preservice teacher self-assessment and reflection and to provide evidence for assessment and accountability (Anderson & DeMeulle, 1998; Wolf & Dietz, 1998). Bannik (2009) piloted a portfolio that included preservice teacher selected videos of teaching samples and reflections. The participant in the case study analysis found the task to be valuable to her development as a teacher. Portfolios can also be constructed as standardized measures (Berrill & Addison, 2010; Porter, Youngs, & Odden, 2001; Tigelaar et al., 2004). Tigelaar and colleagues (2004) document the process of creating a standardized portfolio at one institution in the Netherlands and show that it has the potential to be implemented. However, portfolio use in the United States has been highly contextualized with different universities creating their own portfolio models developed out of specific institutional purposes (Zeichner & Wray, 2001). Portfolios that begin collecting artifacts at the very

beginning of a teacher education program could possibly show preservice teacher learning while also being able to parse out the personal characteristics with which the individual entered the program.

Unfortunately, there is little evidence of portfolio assessments being implemented in a standardized manner that allows for comparisons between teacher education programs. A single, standardized portfolio implemented reliably could produce information that can begin to differentiate between the learning experiences of preservice teachers in different teacher education programs and different specialties within these programs. The fact that a standardized portfolio assessment has not been adopted by many teacher education programs may signal difficulties in the construction or implementation of such a model.

A second attempt at evaluating teacher education is to examine the performance of preservice teachers while teaching in real PK-12 classrooms. This form of assessment has the greatest validity as it is measuring exactly what teacher education programs are attempting to teach. More teacher education programs are emphasizing early-program field experiences that provide preservice teachers the opportunity to teach in PK-12 classrooms (Grossman, 2010). The National Council on Teacher Quality (NTCQ) conducted a study of 134 traditional teacher education programs' student teaching experience. Their results indicate that many teacher education programs have inconsistent evaluation forms within their programs and there was no evidence presented of standardization between teacher education settings (Greenberg et al., 2011). Therefore the current situation of student teaching evaluation neither allows for understanding the

impacts of teacher education on preservice teachers nor the ability to compare across programs.

However, the use of a standardized observation measure, such as CLASS or the Danielson Framework (Danielson, 1996), by more teacher education programs could allow for comparisons across teacher education programs. There is precedence for the implementation of CLASS as an evaluation tool in teacher education (Jamil, Downer, & Pianta, in press; Ripski, LoCasale-Crouch, & Decker, 2011; Wiens, under review). However, there has not been widespread uptake of standardized observation measures in teacher education programs. As Brophy and Good (1986) acknowledge, standardized observation of teachers is resource intensive in terms of training and paying reliable raters to conduct the evaluations. This may explain why more teacher education programs have not adopted these standardized measures. Also, while there are more early-program field experiences, the bulk of classroom teaching remains concentrated at the end teacher education programs (AACTE, 2010).

Another approach to measurement in teacher education is teacher performance assessments (TPA). The most prominent model of this is the edTPA that developed out of the Performance Assessment for California Teachers (PACT; Pecheone & Chung, 2006). The edTPA, was designed "to answer the essential question: 'Is a new teacher ready for the job?'" EdTPA is a standardized measure of preservice teachers' readiness to teach and is currently being piloted in 21 states (edTPA, 2012). The PACT and the edTPA include embedded signature assessments that are created and implemented within individual teacher education programs and a capstone teaching event (Pecheone & Chung, 2006). The edTPA system incorporates evidence of preservice teacher knowledge and

skills including video clips of instruction, lesson plans, student work samples, analysis of student learning, and reflective commentaries (edTPA, 2012). The collected evidence is submitted electronically to a national scoring center that assesses the quality of the preservice teacher's work. Because the measure is so new, there is not yet empirical evidence for its effectiveness.

However, the PACT has been empirically tested and has shown the capability to be implemented across many different California university teacher education programs allowing for comparative analysis. Darling-Hammond (2006) presents data across twelve different teacher education programs suggesting that preservice teachers are most proficient in planning instruction and less consistent in using academic language. A mixed methods case study of elementary teachers (Chung, 2008) as well as an open-ended survey of secondary teachers (Okhremtchouk et al., 2009) both showed that preservice teachers felt that PACT helped them improve their teaching. However, in both studies preservice teachers were concerned with the time burden required to complete the PACT capstone teaching event on top of their other coursework. The PACT is also associated with a financial burden for schools (Guaglianone, Payne, Kinsey, & Chiero, 2009).

The edTPA is a promising standardized measure to compare different teacher education programs. However, it still struggles with similar problems to field experience evaluations using standardized observation measures. Preservice teachers may not have the opportunity to teach enough early in the program to use edTPA to examine teacher learning through the teacher education program. For example the research presented in on PACT was conducted exclusively on the capstone teaching event which occurs in the

final year of the teacher education program (Chung, 2008; Darling-Hammond, 2006; Okhremtchouk et al., 2009; Pecheone & Chung, 2006). While the edTPA provides an authentic measure of teaching ability at the end of the program it does not show growth in teaching knowledge of skills.

A final approach to evaluating teacher education programs has been gaining increasing attention, the use of PK-12 student data to assess the impact of teacher education programs (Plecki, Elfers, & Nakamura, 2012). This procedure involves using VAMs to estimate the effectiveness of an individual teacher. The teacher is then matched to her teacher education program. Then the effects of all the graduates of a particular teacher education program are pooled to estimate the effects of the teacher education program (Gansle, Noell, & Burns, 2012; Goldhaber & Liddle, 2011; Plecki et al., 2012). Gansle, Burns, and Noell (2010) provide a good example of the opportunities and difficulties with using VAMs to judge teacher education program effects in Louisiana. While the analysis is based on student learning—the ultimate goal of teacher education—it is difficult to collect the required data for analysis. Even though the Louisiana Board of Regents is behind the initiative, only nine of twenty-one programs had enough data to calculate the effectiveness of the teacher preparation programs (Gansle et al., 2010). In a related analysis, Gansle and colleagues were forced focused their analysis on teachers in grades four through nine because of the availability of student achievement test data (Gansle et al., 2012).

Furthermore, using VAMs in this way may provide a good indication of the quality of the graduates of different teacher education programs-information that may be useful to school principals trying to hire the best candidate for a position. It is not very

helpful for judging the effectiveness of a teacher education program; however, because it does not account for the preservice teacher characteristics upon entering the program. Therefore, it is impossible to know for certain using VAMs whether the graduates of a particular teacher education program are effective because the program did an excellent job of training them or if it did an excellent job of recruiting individuals who were already good at teaching.

Clearly creating and implementing standardized measurement tools for examining teacher education has many problems. While standardized measures do exist in teacher education they all have problems measuring growth in preservice teachers' knowledge and skills while accounting for the characteristics the individuals already have when they enter the program. Developing measures that can answer these concerns is important to providing information about relative strengths and weakness of different program constructions and individual performances.

Measuring Preservice Teachers' Ability to Recognize Effective Teaching Interactions

As shown previously, research on CLASS has demonstrated the importance of teacher-student interactions to student learning (Gates Foundation, 2012; Pianta et al., 2008). However, due to the lack of teacher experiences early in teacher education programs (Grossman, 2010) it is not possible to use CLASS as a pretest-posttest measure to analyze preservice teacher growth. Instead, measuring preservice teachers' ability to recognize effective teaching interactions may serve as a proxy for their abilities in implementing these teaching behaviors (Jamil, Sabol et al., under review). In a study of more than 400 in-service teachers, Hamre and colleagues (2012) discovered a link between participants' ability to recognize effective teaching interactions and teaching

performance. Situated in empirically and theoretically supported notions of high leverage teacher behaviors, a standardized measure of preservice teachers' ability to detect effective teacher-student interactions that is implemented reliably can produce evidence of teacher knowledge and learning. Combined with data on preservice teacher characteristics, and implemented in a pretest-posttest format, performance on such a measure could demonstrate the impact of the teacher education program and the impact of personal characteristics on teacher learning.

The Video Assessment of Interactions and Learning (VAIL) was developed to assess the abilities of in-service preschool teachers to detect effective teaching interactions in pre-recorded video segments. Based on CLASS, the VAIL is measuring teaching interactions that have been shown to impact student learning (Jamil, Sabol, Hamre, & Pianta, under review). As illustrated in Table 2, the VAIL focuses on three specific dimensions of the CLASS framework: regard for student perspectives, instructional learning formats, and quality of feedback. When combined these three dimensions of teaching interactions can demonstrate a preservice teachers' ability to detect effective teaching interactions.

The VAIL is a relatively new measure, but it has been implemented with both in-service and preservice teachers. Hamre et al. (2012) first used the VAIL as an outcome measure in a study of 440 early childhood teachers. In their study, Hamre and colleagues randomly assigned half the participants to participate in a course designed to help them improve their teacher-student interactions. Treatment and control groups were given a battery of surveys, had their teaching performance rated by the CLASS Pre-K measure, and completed a two-video version of the VAIL. Participants in the treatment group

Table 2
CLASS Dimensions Measured by the Video Assessment of Interactions and Learning

Domain	Emotional Support	Classroom Organization	Instructional Support
Dimension	Regard for Student Perspectives	Instructional Learning Formats	Quality of Feedback
Indicators	• Teacher flexibility • Student autonomy • Student Expression	• Variety of learning format • Promotion of student interests • Clarity and engaging approach	• Feedback loops • Encouragement of responses • Expansion of performance
Support	Pianta & Hamre, 2009; Roeser, Eccles, & Sameroff, 2000; Stefanou, Perencevich, DiCinto, & Turner, 2004	Brophy & Good, 1986; National Research Council, 2000; Pianta & Hamre, 2009	Hamre & Pianta, 2007; Taylor, Pearson, Peterson, & Rodriguez; 2003; van de Pol, Volman, & Beishuizen, 2010

showed more effective teacher-student interactions and also showed greater skills in detecting effective teacher-student interactions as measured by VAIL (Hamre et al., 2012). Jamil, Sabol, and colleagues (under review) examined a set of 270 preschool teachers drawn from the larger Hamre et al. (2012) study who had not experienced the intervention. Jamil and colleagues found that teaching experience was the only demographic variable related to performance on the VAIL. They also found performance on the VAIL to predict teaching interactions in the instructional support domain. Because the VAIL bases its validity on CLASS domains, these findings linking performance on the two measures is significant. This connection provides evidence that VAIL is measuring a meaningful skill for effective teaching.

The VAIL follows the assumption of CLASS that teachers make sense of effective teaching in similar ways across teaching specialties allowing for a single measure for all teachers (Pianta & Hamre, 2009a). CLASS draws from theory and research in child development and parenting as well as classroom interactions (Hamre & Pianta, 2007). Additionally, the CLASS instrument has been used repeatedly in classrooms across grade levels and teaching specialties (Allen, Pianta, Gregory, Mikami, & Lun, 2011; Grossman et al., 2010; Pianta, Belsky, Houts, & Morrison, 2007). Analysis of the implementation of the VAIL with preservice teachers from different teaching specialties found that teaching specialty was not a significant predictor of participants' ability to detect effective teaching interactions (Wiens & Hessberg, 2011).

One advantage of the VAIL is its ability to be administered at several time points within a teacher education program. Wiens and colleagues (2013) describe the implementation of VAIL at three time points in a five year bachelor's plus master's degree program and in each year of a two year postgraduate master's degree teacher education program. This makes the VAIL a potentially appropriate standardized measure for evaluating the effects of a teacher education program. Because the VAIL has already been implemented in multiple settings, there is also evidence that it can be used for comparisons between teacher education programs.

Limitations in Measurement of Teacher Education

The need for standardized measures of teacher education is great, but there are few viable options for teacher education programs and policy makers. While there are examples of standardized measures in teacher education each has its own specific limitations. The complexity of teaching requires multiple measures that can be

implemented in a variety of teacher education settings. This purpose of this study is to examine one teacher education program using a video-based standardized measure of preservice teachers' ability to detect effective teaching interactions.

After reviewing the literature, questions remain about the impacts of teacher education programs on preservice teachers due to the lack of good measurement tools. The purpose of this study is to examine the impacts of a teacher education program and personal characteristics on preservice teacher knowledge and learning. Through an analysis of teacher and program characteristics and ability at entry to the program, this study seeks to identify if preservice teachers are better able to identify effective teacher interactions at program exit. It also seeks to understand if the ability to detect effective teaching interactions is predictive of preservice teachers' ability to perform these high impact teaching behaviors.

CHAPTER THREE

METHODS

This chapter explains the methods used to address the three research questions that guide this study. This methods chapter will begin with an overview of the study. Following the study overview, the study setting and participants will be described. Next, each of the measures employed in this study will be explained. Finally the analysis strategy used to specifically address each of the research questions will be provided.

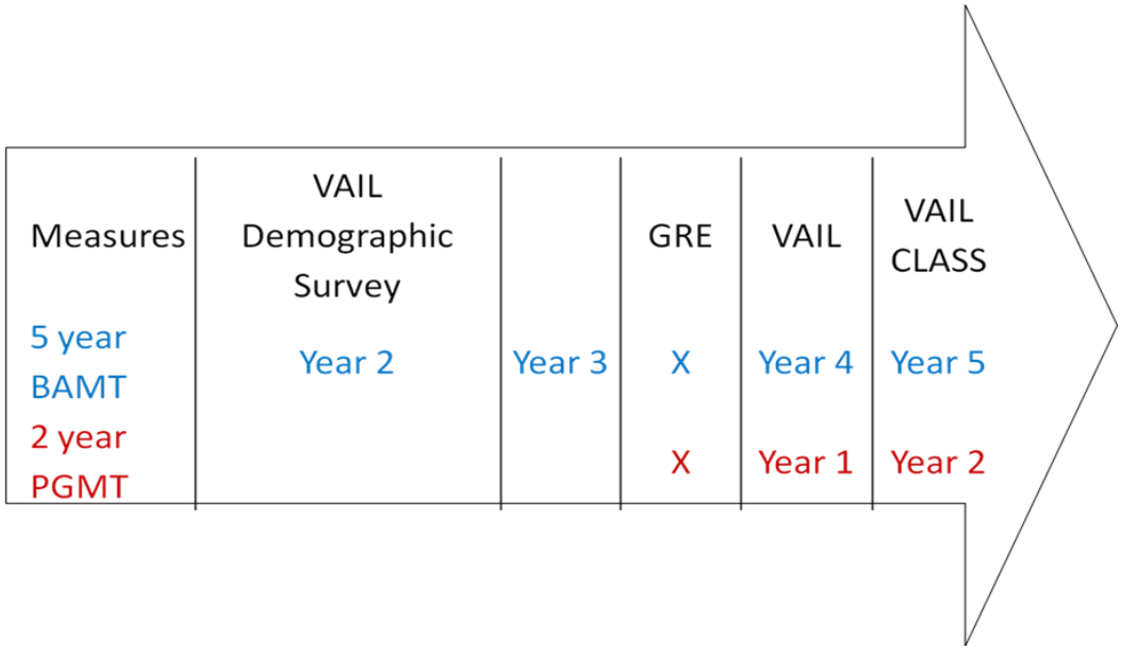
Study Overview

Preservice teachers and individuals enrolled in an introductory teacher education course in a mid-sized teacher education program at a mid-Atlantic, public university were participants in this study. The teacher education program consists of a five-year bachelor's plus master's degree (BAMT) and a two-year post-graduate master's degree (PGMT). Both programs, combined with passing state required exams, lead to state teaching certification. The two programs are integrated whereby the BAMT and PGMT students take classes and participate in field experiences together.

The VAIL is administered at three different points in the teacher education program as illustrated in Figure 2. First, the VAIL is administered to students enrolled in an introduction to education class which is traditionally taken by prospective teachers in the second year of their five year teacher education program (PGMT students do not take

this course). The introductory course is not a methods course, but instead the curriculum is designed as a survey of American education where students study the historical, philosophical, political, and legal aspects of the educational system in an effort to study what it means to be a teacher. Many, but not all, of the students who take this course go on to enroll in the teacher education program. BAMT students complete the VAIL again in the spring semester of their fourth year following three semesters of methods courses. The PGMT students take the VAIL for the first time in the spring of their first year of the program. The final administration of the VAIL for both programs is in the spring semester following the fall one-semester student teaching experience.

Figure 2
Timeline of Measure Implementation



Observed student teaching performance is also collected using the CLASS system. The one-semester student teaching experience occurs during the fall semester of the final year of both the five-year and post-graduate programs. The preservice teachers were

instructed to video-record a teaching lesson during a two-week time frame towards the end of the student teaching semester when they were assumed to have full teaching responsibilities. CLASS-trained coders then analyzed the video-recordings and recorded the results in a teacher education database. Many of the CLASS coders were university supervisors responsible for mentoring the student teachers and provided codes on their own students.

Demographic information including race, gender, and admission testing data as well as program data are collected by a combination of surveys and official student records kept in the Teacher Education Office. Program data include teaching specialty (English, Elementary, etc.) as well as BAMT, PGMT, or non-teacher education student enrollment status.

The teacher education department first implemented the VAIL in the spring semester of the 2009-2010 academic year. This study analyzes data from three years of VAIL data including the 2009-2010 academic year through the 2011-2012 academic year. CLASS and demographic data were also used from the same academic years. These demographic data together with the VAIL and CLASS data described here allow analyses that address this study's research questions. The participants, setting, measures, and analysis are described in detail below.

Setting

Data in this study were collected at a mid-sized public university. The university is considered a highly selective national university by U.S. News and World Report (U.S. News, 2012). In total, the teacher education program graduates per year approximately 130 students in teaching specialties including elementary education, English/language

arts, mathematics, science, social studies, foreign language, special education, and health/physical education.

Participants

Data for this study include all participants who have taken the VAIL during the three years of VAIL administration (academic years 2009-2010 through 2011-2012) either once, twice, or three times. The complete dataset includes 787 participants and is described in Table 3. Of the complete dataset, 30.6% are not enrolled in the teacher education program, 15.5% are post graduate master's degree students (PGMTs), and 53.9% are enrolled in the five-year bachelors plus master's degree program (BAMTs). The participants are 79.3% female and 19.9% male with 1.0% missing data. The participants describe themselves as 72.6% Caucasian, 5.0% Asian, 5.5% African American, 1.7% Hispanic, and 15.1% other or not specified.

Measures

Predictor Variables

Demographic data were taken from two surveys and administrative records. During the administration of the VAIL in the introductory course, participants also completed a survey including questions regarding race, ethnicity, program of study, and high school characteristics. Teacher education office administrative records also provided demographic and entrance test score data. If there were disagreements between the different datasets, the administrative records were used.

Gender. Gender was also included as a dummy variable.

Race. Racial identification was included using the following categories: Caucasian, African American, Asian, and other/not specified.

Table 3
Participant Characteristics

Group	N	%
Gender		
Male	154	19.9
Female	623	79.3
Missing	10	1.0
Ethnicity		
African American	43	5.5
Asian	39	5.0
Hispanic	13	1.7
Caucasian	571	72.6
Other	64	8.1
Not Specified	55	7.0
Teaching Specialty		
Elementary	191	24.3
English	62	7.9
Foreign Language	48	6.1
Health/PE	18	2.3
Math	27	3.4
Science	33	4.2
Special Education	92	11.7
Social Studies	71	9.0
Not enrolled in TEd.	241	30.6
Teacher Education Program		
BAMT	424	53.9
PGMT	122	15.5
Not enrolled in TEd.	241	30.6
Total	787	100.0

Academic Aptitude. The Graduate Record Exam (GRE) served as a proxy for general academic aptitude. The GRE (www.ets.org/gre) is required of all students for entrance to the graduate program in the teacher education program. Therefore, PGMT students have to take the GRE prior to admission. The BAMT students are required to take the GRE prior to advancing to the fifth year of the five-year program. The teacher education program sets minimum GRE score requirements for entrance although exceptions can be made on occasion for students in the BAMT program who present scores that are close to the cutoff point and are otherwise making satisfactory academic progress. This study used all three of the GRE tests including verbal reasoning, quantitative reasoning, and analytical writing on the 200-800 point scale (used prior to 2011). Any GRE data collected in the newer scaled scores were converted to the older scale using a table provided by the Educational Testing Service.

Teaching Specialty. Teaching specialty was also included with the following categories: pre-k/elementary, English/language arts, foreign language, health/physical education, math, science, special education, and social studies.

Teacher Education Program. When applicable, this study used program information including a variable where the options are "not enrolled in teacher education", "BAMT", or "PGMT".

Outcome Measures

Video Assessment of Interactions and Learning. The VAIL was used to determine preservice teachers' ability to identify effective teaching strategies in video segments of real classroom teaching environments. The participants watched three short videos (2-3) minutes. The VAIL was originally developed for a large-scale study of

preschool teachers (Hamre et al., 2012), so each video features a preschool language arts lesson taught by a veteran in-service teacher that demonstrated effective teaching characteristics as measured by CLASS. The selected videos represent a dimension within each of the three CLASS domains: quality of feedback in the instructional support domain, instructional learning formats in the classroom organization domain, and regard for student perspectives in the emotional support domain.

After watching the video, participants had the opportunity to provide five effective teaching strategies they identified from the video in an open-ended format. However, only one attempt at a strategy-example pair was required to advance to the next video. A strategy is a general marker of effective teaching. Examples of effective teaching strategies included in the VAIL would be scaffolding, eliciting student ideas, and variety of instructional modalities. For each strategy, the participant had the opportunity to provide a specific example of the strategy taken from the video. The assessment defines an example as, "A teaching method used to meet a specific goal". In other words, examples constituted specific actions observed in the video. For example, if a participant noted scaffolding as a strategy a matching example might consist of the teacher helping the student sound out the word the student was struggling to read.

Videos were selected based on their ability to provide examples of effective teaching strategies in the chosen CLASS dimensions. A prompt was included with each video to give participants direction for what to look for in the video. The prompts for the three videos were as follows:

1. Name up to 5 strategies the teacher is using to **show she values children's ideas and points of view and encourages children's responsibility and independence.**
2. Name up to 5 strategies the teacher is using **to engage the students in the lesson and hold their attention.**
3. Name up to 5 strategies the teacher uses to **effectively provide feedback and extend students' learning, skills, and persistence.**

Responses supplied by participants were open-ended and were coded for accuracy against a master code list created by master coders who were also CLASS trained and helped to create the VAIL and then reconciling differences through discussions that were based on standards identified in the CLASS (VAIL 2010). The VAIL uses a standardized rating description as outlined in the VAIL Coding Manual (2010) which was used to guide all coding decisions. The VAIL was designed so that CLASS-specific terminology was not necessary to perform well on the assessment. The master codes were descriptive of aspects of effective teaching and the language was not tied to the CLASS system.

As described in Wiens and colleagues (2013), when a CLASS-matched strategy was identified, a breadth score was also assigned to identify the specific indicators within the CLASS dimension identified by the participant. For example, for one of the videos, the CLASS dimension was Instructional Learning Formats (ILF—a part of the Classroom Organization domain); which has the following indicators: (1) engaging approach, (2) variety of modalities, (3) student interest, and (4) clarity of learning objectives. A CLASS-matched strategy for this video would have to match, or constitute a reasonable synonym of one of these four indicators. For example, a participant can submit

"effective questions" as a strategy under (1) engaging approach. This would be credited as a CLASS-matched strategy.

The number of unique indicators supplied by participants was then summed to create a breadth score for the entire set of responses for that video. In the ILF/Classroom Organization example from above, a participant may complete four strategy-example pairs. This participant would have been coded as supplying the following strategies:

Pair 1: engaging approach

Pair 2: none

Pair 3: student interest

Pair 4: engaging approach

This participant would receive a two for the breadth score under ILF because he or she only provided two different strategy types (engaging approach and student interest).

Additionally, if both the strategy and example supplied were correct, the response was coded based on whether the example was an accurate example of the strategy identified. Thus, if the participant was looking for strategies based on ILF, he or she may have identified effective facilitation. If the example the participant provided was an example of ILF, he or she would get credit for as an example. And if the example was an example of effective facilitation the participant would also get credit for a match. However, it was possible for a participant to identify an accurate example while not matching that example to its strategy pair. In this case the participant would get credit for an example and strategy, but no credit for a strategy-example match.

Table 4 demonstrates how a participant's responses may be coded. In the first pair the participant would be credited for providing a correct strategy and a correct

example. The participant would also be coded as providing a strategy-example match as both the strategy and the example belong to the same CLASS indicators. The second strategy-example pair would also be coded as correct for both strategy and example. However, the second pair would not be coded as providing a strategy-example match because they come from different CLASS indicators. This response set would also be coded as a breadth score of one because both supplied strategies belong to the same CLASS indicator.

Table 4
VAIL Coding Example

	Participant Response	Code
Strategy 1	Teacher asks good questions	Engaging approach
Example 1	While one student is volunteering the teacher asks the rest of the class to help her	Engaging approach
Strategy 2	Teacher asks effective questions	Engaging approach
Example 2	Students have the opportunity to move around the room as part of a word find	Variety of modalities

The completion score measured how many responses the participants wrote for each video. Participants were coded for each attempt at identifying a strategy and example even if the strategy and example were not correctly identified. Each participant was required to provide at least one strategy and example to continue in the assessment. While there was the opportunity to identify five strategies and examples, only one response was required to continue with the assessment. Any strategy-example pairs that were left blank were coded as a zero.

VAIL Scoring. Research assistants who were doctoral students in the Education School with which the teacher education was associated participated in a half-day training session that included an extended introduction to VAIL coding. The training sessions culminated in a reliability test where research assistants coded two sets of responses and were considered reliable if their codes showed exact agreement with the master codes at least 80% of the time. All research assistants completed reliability tests every one to two weeks while they were coding. Each time the assistant needed to score at least an 80% exact agreement. If an assistant failed to make 80% agreement, he or she stopped coding, retrained, and passed another reliability test before coding again. The coding team showed strong reliability demonstrating a minimum of 80% exact agreement across the three years of data on the 20% of the VAIL responses that were double coded. Cohen's Kappa was also calculated between raters with an alpha of .70 in the first year, .76 in the second year, and .76 in the third year of data. These alpha values indicate an acceptable level of inter rater agreement (Landis & Kock, 1977).

To analyze the VAIL data, sum scores were calculated. Previous analysis of VAIL data with in-service teachers presented evidence to support using a one-factor model for compositing VAIL scores using the strategy, example, match and breadth scores (Jamil, Sabol et al., under review). The completion variable is analyzed separately because it does not conceptually measure a participant's ability to detect effective teaching interactions; instead it measures participants' persistence in completing the assessment.

Jamil, Sabol, and colleagues (under review) suggest an analysis strategy that standardizes values within the different videos and then composites the videos into a

single score. However, it may be easier to understand the results of the VAIL, particularly when examining longitudinal change, using a sum score. Use of a sum score instead of a composited mean score allows for a clearer interpretation by providing a raw number that can be compared easily across program years. Additionally, using a sum score also facilitates comparison of participant scores across contexts and administrations of the VAIL by providing a fixed number for the final score. The drawback of this approach is that the videos do not all have the same total possible points and therefore one video might have a slightly smaller weight in the overall score than the other videos. The total possible points for the Regard for Student Perspectives video is 19, Instructional Learning Formats is 19 as well, and the Quality of Feedback video total is 20. The differences in possible points comes from the breadth score which has a maximum of four strategies in Regard for Student Perspectives and Instructional Learning Formats, while there are five total strategies in Quality of Feedback. While a sum score makes the Quality of Feedback video slightly more important, the benefits of a sum score outweigh this disadvantage.

The validity of the VAIL is based on the strong empirical support behind the CLASS framework described below. As a relatively new measure, the VAIL is still developing an empirical base for its validity and reliability. Performance on the VAIL has shown a relationship to teaching performance measured by CLASS in in-service preschool settings (Hamre et al, 2012; Jamil, Sabol et al., under review). It has also shown adequate reliability data with Hamre and colleagues reporting an 82.5% exact match in the 20% of double-coded data. In a smaller sample taken from the larger Hamre et al. study, Jamil, Sabol, and colleagues reported an overall *Kappa* value of .567.

The VAIL has also been used in a teacher education setting demonstrating that it can be administered reliably at several points in a teacher education program (Wiens et al., 2013). Wiens and colleagues (2013) reported a reliability *Kappa* statistic as .70. Additional analysis of the preliminary VAIL data indicated that there were no significant differences in performance on the VAIL among the different teaching specialties (Wiens & Hessberg, 2011). VAIL data has not yet been examined in a longitudinal format in a teacher education setting nor has it been correlated with teaching performance as measured by CLASS in a teacher education setting.

Analysis in this study used VAIL sum scores derived from different points in the teacher education program. The nature of the available VAIL data as well as the research questions required the use of four different VAIL scores taken from different points in the teacher education program. Each VAIL score was pulled to address the specific research questions and is described in detail below.

Entry VAIL. The first VAIL sum score is the entry VAIL. This variable includes all participants who took the VAIL in the introductory course or, for the two-year PGMT students, in the first year of their program. The entry VAIL functions as a pre-test measure and approximates preservice teacher ability to detect effective teaching interactions prior to receiving training from the teacher education program.

Initial and Final VAIL. The second set of VAIL variables are necessitated by the nature of the currently available VAIL data and are designed for longitudinal analysis. Ideally, the simplest format for conducting longitudinal analysis on VAIL data would be to use the scores at entry as described in the previous paragraph and then the VAIL scores at the end of the program. However, currently only three years of data have been

collected and therefore participants who have completed the VAIL in the introductory course have not yet completed the teacher education program. Analysis of the data shows that only 20 participants have completed the VAIL during the introductory course and have completed the VAIL in the final year of their program. This number is too small for meaningful statistical analysis.

An alternative strategy was used with the currently available data. Analysis of the data identified 132 participants who have completed the VAIL more than one time. From these cases, two VAIL scores were developed. The first is the initial VAIL score. The initial VAIL score pertains to the first score for participants that have taken the VAIL more than once. Of these initial VAIL scores, 16.3% were in the second year of the BAMT program, 10.6% were from the third year of the BAMT program, 52.8% were from the fourth year of the BAMT program, and 20.3% were from the first year of the PGMT program. The final VAIL is the companion score to the initial VAIL. The final VAIL is the last VAIL score a participant has in the data set. Of the final VAIL, 3.3% were in the third year of the BAMT program, 15.4% were in the fourth year of the BAMT program, 1.6% were in the first year of the PGMT program, and 21.1% were in the second and final year of the PGMT program. This data shows that the both the initial and final VAIL scores are unique and different from the other VAIL scores described in this study.

Exit VAIL. Finally, an exit VAIL was created. The exit VAIL variable included the VAIL scores from the final year of an individual's teacher education program. For some participants the exit VAIL might be the only time they took the VAIL because the assessment was introduced and administered at all identified time points in the program

from the beginning of implementation. In other words, in the 2009-2010 year when the VAIL was first implemented it was given to all students in the designated positions in their teacher education programs. However, all participants who have exit VAIL data have completed their student teaching and have had equal access to their respective teacher education programs. The exit VAIL comes from those preservice teachers who have completed their student teaching experience and therefore many also have CLASS data.

Classroom Assessment Scoring System. CLASS has been utilized by researchers as an effective measurement in elementary and secondary classrooms both in the United States (Graue, Rauscher, & Shefinski, 2009; La Paro et al., 2009; Malmberg & Hagger, 2009) and internationally (Cadima, Leal, & Burchinal, 2010; Parkarinen et al., 2010). CLASS-based studies consistently find associations between observable classroom behaviors outlined in the CLASS protocol and student development and learning. For example, in a longitudinal study of 147 kindergartners through first grade, Curby, Rimm-Kaufman, and Ponitz (2009) found that teachers high in emotional support had students that demonstrated faster growth in phonological awareness. In another study, Pianta, and colleagues (2008) examined nearly 800 students in various elementary classrooms and found a link between emotional support and reading achievement. Additionally, the Measures of Effective Teaching (MET) Project assessed nearly 3000 teachers and found a positive relationship between teachers' ratings on CLASS and study value-added estimates across grades K-12 and subject matter (Gates Foundation, 2012). Consequently, several recognized educational research agencies such as The Gates Foundation, Educational Testing Service, and the United States National Institute of Child Health and

Human Development have included CLASS as part of their in-service teacher studies (Gates Foundation, 2010; Ewing, 2008).

In the program under study, preservice teachers in their final year complete a one-semester student teaching placement in the fall semester. The preservice teachers video-record themselves during a specified period of time when they are assumed to have taken on full teaching responsibilities. From the videos two sets of CLASS codes are generated by trained raters that are then composited into one mean score. Raters were initially trained to reliability on the tool through a rigorous two day training session where they learned the CLASS framework and conducted multiple practice tests. Next, observers passed a reliability test, using the CLASS tool successfully across multiple classroom situations. All raters must demonstrate an 80% agreement of within one score of a master coding list to be considered reliable. In this study, the raters were often the university supervisors assigned to mentor and evaluate the student teachers.

Analysis of the CLASS is generally conducted at the domain level. Therefore, the dimensions are composited into their three respective domains. Each domain is then entered separately into the analysis as described in the following section. In this sample, participants were coded twice during their student teaching using the CLASS. This was sometimes during the same teaching lesson or during separate lessons. These codes were averaged together to create one set of codes for each individual. Scores were then composited. In the three cohorts of preservice teachers with CLASS data the reliability of compositing scores varied. For the emotional support domain, the reliability coefficient was α=.684. For the classroom organization domain, reliability coefficient was α=.628. Finally, the instructional support domain produced a reliability coefficient

of α=.710. Overall, the reliability coefficients were sufficiently high to analyze the CLASS observation data at the domain level.

Analysis

Analysis in this study occurred in three stages to address the three research questions. All analyses described in this section were conducted using the PASW 18 statistical package. Preliminary descriptive statistics of all variables described in this section were also conducted as well as correlation analyses. Results of the analysis are described in Chapter 4.

Research Question One

The first research question seeks to examine the association between preservice teachers' individual demographic characteristics, teaching specialty, and teacher education program on their ability to detect effective teaching interactions. To answer this question, two analyses were conducted. The first analysis examined the differences between those students who took the VAIL in the introductory course and went on to enroll in the teacher education course and those that did not. An analysis of variance (ANOVA) was conducted to determine if there was any significant difference between students that enrolled in teacher education and those that did not. The dependent variable in this analysis was the entry VAIL scores.

The second analysis implemented a regression equation. The dependent variable in this analysis was the entry VAIL score as described previously. The predictor variables included the following demographic variables: gender, race, and GRE score. The gender variable was dummy coded male=1, female=0. Race was also dummy coded with Caucasian being the comparison because the majority of participants were

Caucasian, and the following groups entered as dummy-coded variables: African American, Asian, Other/Not Specified. GRE data were entered separately as raw scores for verbal, quantitative and analytic writing.

Additional predictor variables included teaching specialty and program. The teaching specialty variable was dummy coded as well for the regression analysis with the elementary group serving as the comparison because the elementary program is the largest program and there is little overlap in coursework between the elementary program and other programs. The following groups were created as dummy codes: English/language arts, foreign language, health/physical education, mathematics, science, special education, and social studies. The program variable included BAMT as the comparison group with PGMT and non-Teacher Education Student (Non-TEd) included as dummy coded variables.

As described in Chapter 4, analysis results were interpreted for the predictive value of the overall regression equation as well as the individual predictors and their ability to explain the variance in the entry VAIL score.

Research Question Two

The second research question sought to examine if, after controlling for initial skill, do preservice teachers' individual characteristics, teaching specialty or teacher education program predict change in their ability to identify effective teaching interactions. Prior to estimating the entire regression equation, an initial regression was conducted to understand the relationship between the initial and final VAIL scores.

Next, to answer this research question the initial VAIL and final VAIL variables were used to conduct a linear regression. In order to control for the initial skill in

detecting interactions, the first VAIL score was entered as the first predictor variable. Other predictor variables mirrored those used in answering question one: gender, race, GREQ, GREV, GREA, program, and teaching specialty.

Also described in Chapter 4, results of this analysis indicated the relationship of initial skill and later skill in detecting interactions and whether the personal and programmatic characteristics remained influential.

Research Question 3

The third question sought to understand the association between performance on the VAIL and student teaching performance. To answer this question three different regression equations were estimated. Each domain of the CLASS (emotional support, classroom organization, and instructional support) served as dependent variables in the three separate regressions. Again, linear regressions were conducted to estimate the association between the two variables. The addition of the previously described predictor variables allowed for greater precision in isolating the shared variance between the VAIL performance and CLASS performance. The first variable entered was the exit VAIL score followed by individual, teaching specialty, and program variables as described in the previous two analyses.

Inspection of the available data indicates that a quarter (25.4%) of the preservice teachers in the exit VAIL data set were missing academic aptitude data (GRE). Using list wise deletion, the analysis dropped these individuals out of the estimation. Therefore, analysis of variance (ANOVA) was also computed to determine if there were significant differences between participants with aptitude scores and those without. These ANOVAs were conducted for both the exit VAIL variable and the three CLASS domain scores.

Conducting regressions on the results of the preservice teachers' attempts to detect effective interactions provided a framework for answering the research questions for this study. Analysis of individual and program data, VAIL data, and CLASS data provide an examination of the use of a standardized measure in a teacher education program.

CHAPTER FOUR

RESULTS

This chapter explains the results of the analyses described in Chapter 3. The first part of this chapter provides preliminary analyses describing the data for illustrative purposes and the second portion describes the results for each research question. The first question examined the predictive ability of gender, race, academic aptitude, teaching specialty, and teacher education program on participants' ability to detect effective teaching interactions at program entry. The following section discusses the results of analysis conducted to determine if these demographic and program characteristics remained significant on a final test of the ability to detect teaching interactions after taking into account prior ability. The final section examines the results of analyses conducted to examine the association between the ability to detect effective interactions and observed teaching performance.

Preliminary Analysis

A series of preliminary analyses were conducted to describe the data. This builds on the information already presented in Chapter 3 which includes a description of the participants and explains the scoring and variable construction of the instrument. Basic descriptive statistics for scale variables are presented in Table 5. The VAIL scores at

entry were M=15.78, *SD*=6.56. The initial VAIL variable pertains only to those participants who took the VAIL more than once and represents the first of these multiple administrations of the VAIL (*M*=16.01, *SD*=7.03). The final VAIL is the last of these multiple administrations of the VAIL (*M*=15.58, *SD*=7.31). The large range in minimum and maximum scores combined with the large standard deviation shows a substantial amount of variance among individual performance on the VAIL. This is echoed by the exit VAIL data (*M*=16.24, *SD*=7.63). Visual examination of histograms for these variables also indicated that the variance in performance on the VAIL was normally distributed across the preservice teachers. The change in VAIL score (ΔVAIL) was computed by subtracting the initial VAIL from the final VAIL. Analysis of the data indicated that an equal amount of individuals scored higher (62) on their second VAIL as individuals who scored lower (62) on their second VAIL with seven participants showing no change between times.

Table 5 also includes descriptive statistics for the GRE scores of the sample. CLASS scores are also included in Table 5 and are described at the domain level including emotional support, classroom organization, and instructional support.

Analysis of the bivariate correlations described in Table 6, demonstrate an overall lack of significant relationships between variables with two notable exceptions—both pertaining to the exit VAIL. First, there was a significant relationship between race and scores on the exit VAIL. This result indicates that, for this sample, the participants' race was correlated with VAIL performance in the final year of their teacher education program (*r*=.209, *p*=.006). Otherwise gender, being enrolled in teacher education, GRE

performance, and observed classroom teaching were not significantly correlated with performance on any of the VAIL variables.

The second significant finding illustrated in Table 6 is a positive correlation between the exit VAIL and two of the domains of the CLASS measure. Participants that had already completed their student teaching showed a significant relationship between the exit VAIL and teaching performance in the emotional support domain (r=.183, *p*=.037). These preservice teachers also showed a significant correlation between exit VAIL and the instructional support domain (r=.176, *p*=.046). Administrations of the VAIL prior to student teaching did not produce any significant correlations with the CLASS measure.

Analyses of Detecting and Implementing Teaching Interactions

In this section the results of analyses conducted to answer each research question will be described. The statistical procedures will be explained with results included for each analysis. The descriptions will proceed in the order of the research questions.

Predicting Initial Ability to Detect Effective Teaching Interactions

The first research question in this study sought to examine if gender, race, academic aptitude, teaching specialty or teacher education program of preservice teachers predict performance at entry on the VAIL. Analysis of this question included two parts which confirm correlation analysis as described above. In the first part, an analysis of variance (ANOVA) was conducted to determine if there were mean differences in students who took the VAIL in the introductory course between students who went on to enroll in the teacher education program and those that did not. ANOVA results of 381 VAIL between these two groups (did not enroll, *M*=15.52; did enroll,

Table 5
Descriptive Statistics

Variable	N	Mean	Standard Deviation	Min.	Max
Entry VAIL	456	15.78	6.56	0	34
Initial VAIL	132	15.98	7.67	0	37
Final VAIL	132	15.58	7.31	0	32
ΔVAIL	132	-.13	9.60	-21	24
Exit VAIL	185	16.24	7.63	0	42
GRE-V	355	542.79	83.98	330	800
GRE-Q	356	635.42	92.20	310	800
GRE-W	351	4.41	.60	3	6
CLASS-ES	176	5.33	.63	3.88	7.00
CLASS-CO	176	4.91	.72	3.00	6.67
CLASS-IS	176	3.74	.96	1.83	6.33

Table 6
Bivariate Correlations with VAIL results

	Entry VAIL	N	Initial VAIL	N	Final VAIL	N	Δ VAIL	N	Exit VAIL	N
Male	.009	456	-.165	131	.040	131	.150	131	.052	185
Race	.045	456	-.013	127	-.015	127	-.037	127	.268**	181
GRE- V	.096	135	-.072	93	-.048	93	.003	93	-.044	138
GRE- Q	.113	135	-.043	94	.071	94	.149	94	.021	139
GRE-W	.060	133	-.102	91	.084	91	.147	91	.144	136
Teaching Specialty	.036	456	-.155	132	-.159	132	.022	132	-.118	185
TED Program	.028	456	.070	132	-.052	132	-.102	132	-.060	185
CLASS-ES	.119	44	.025	53	.156	53	.037	53	.183*	130
CLASS-CO	.051	44	-.151	53	.108	53	.005	53	.097	130
CLASS-IS	.082	44	-.082	53	.007	53	.121	53	.176*	130

*p<.05
**p<.01

participants indicated that there was no significant difference in performance on the *M*=15.90, *p*=.58) prior to enrolling in the teacher education program. The lack of significant difference between these groups allows for the exclusion of non-teacher educations students from the remaining analyses. Results of this analysis indicate that participants' ability to detect effective interactions did not differ prior to program entry between students who would go on to enroll in the teacher education program and those who would not.

To conduct the second part of this analysis to address if gender, race, academic aptitude, teaching specialty, and teacher education program can predict the ability to detect effective interactions, the entry VAIL score was used as an outcome variable. Demographic and programmatic variables were then entered as predictors and the results are described in Table 7.

For each set of dummy coded variables (race, teaching specialty, and teacher education program) one variable was left out of the analysis to serve as a comparison group. Therefore the results reported in Table 7 show the performance relative to the comparison group. Beta values reported here are the standardized values of the predictors on the entry VAIL score.

In this regression equation, the overall model is not significant (adjusted R^2= -.016). Because the entire model was not significant, it is important not to emphasize individual predictors because the individual relationships are not entirely clear. In this analysis, individual demographic variables including gender and race did not show a significant relationship with entry VAIL. Likewise, programmatic variables were also

Table 7
Standardized Regression Coefficients: Entry VAIL Score

Total N=133	
Predictors	ß
Male	-.164
Race (Comparison=Caucasian)	
African American	.072
Asian	.096
Hispanic	-.122
Other/Not Specified	.125
Academic Aptitude	
GRE Verbal	.041
GRE Quantitative	.163
GRE Analytical Writing	.103
Teaching Specialty (Comparison=Elementary)	
English	-.091
Foreign Language	.009
Health/PE	-.113
Math	-.088
Science	-.011
Special Education	-.032
Social Studies	.094
Teacher Education Program (Comparison=BAMT)	
PGMT	.032
Final adjusted R^2	-.016

examined in this model. Teaching specialty and teacher education program were not significant predictors of performance on the VAIL at program entry.

Results of the analysis described in this section as well as correlation analysis indicate that there was no significant difference between those participants who later enrolled in the teacher education program and those that did not in their ability to detect effective teaching interactions while still in the introductory course. Additionally, gender, race, academic aptitude, teaching specialty, and teacher education program do not predict preservice teachers' ability to detect effective teaching interactions at program entry.

Controlling For Initial Ability and Predicting Final Ability to Detect Interactions

The second research question regarded the predictive quality of personal and programmatic characteristics on the ability to detect effective teaching interactions after controlling for an initial measure of that ability. Two separate regression equations were conducted to address this question.

In order to conduct this analysis, it was first important to understand the relationship between the first administration of the VAIL and the final score on the VAIL. The resulting regression equation showed a significant relationship between initial ability and final ability (R=.176, p=.043). Therefore, initial performance on the VAIL was predictive of later performance on the same measure.

To address the research question, a regression equation was estimated building on the first equation. Again the final VAIL variable was used as the outcome variable. In this equation, the initial VAIL score was entered as the first predictor followed by the same personal and programmatic characteristics described in the previous research question. The results of this analysis are illustrated in Table 8.

Table 8
Standardized Regression Coefficients: Final VAIL Score

Total N=90	
Predictors df(17,72)	ß
Initial VAIL Score	.171
Male	.327*
Race (Comparison=Caucasian)	
African American	.005
Asian	-.022
Hispanic	.085
Other/Not Specified	-.075
Academic Aptitude	
GRE Verbal	-.131
GRE Quantitative	.005
GRE Analytical Writing	.170
Teaching Specialty (Comparison=Elementary)	
English	.004
Foreign Language	-.202
Health/PE	-.165
Math	.065
Science	-.226
Special Education	-.067
Social Studies	-.234
Teacher Education Program (Comparison=BAMT)	
PGMT	.039
Final adjusted R^2	-.023

*p<.05

The overall model was also not significant (adjusted R^2= -.023, *p*=.594). When all the variables were added to the equation, the initial VAIL score did not predict performance on the final administration for preservice teachers (ß=.171, *p*=.152). Therefore, the other predictive variables are mediating the predictive power of the initial VAIL performance.

Gender was the lone significant predictor of performance on the final VAIL score (ß=.327, *p*=.056) at conventional levels of statistical significance. In this equation female participants performed significantly better on the final VAIL than their male counterparts even after performance on the initial VAIL score was taken into account. Race and GRE scores provided no significant predictive value in this equation. With elementary preservice teachers as the comparison group there was variation in performance of different teaching specialties. Science preservice teachers showed lower final VAIL scores than elementary teachers but only by a less stringent statistically significant threshold (ß= -.226, *p*=.093). There were also no significant predictive differences between BAMT and PGMT preservice teachers.

Regression analysis supports the non-significant findings from correlation analysis. There were no significant bivariate relationships between gender, race, GRE score, teaching specialty, or teacher education program and initial or final VAIL score. Additional analysis examined if there was a relationship between these variables and the change in VAIL score. Bivariate correlation analysis found no significant relationships.

Analysis described in this section showed a significant relationship between participants' initial ability to detect teaching interactions and their abilities to do the same later in the teacher education program. However, as an entire model, gender, race,

academic aptitude, teaching specialty, and teacher education program were not significant predictors of participants' abilities to detect effective teaching interactions the second time this ability was assessed. While prior ability to detect effective teaching interactions does predict later ability on the same measure, this relationship did not remain when additional variables were included.

Detecting Interactions and Observed Teaching Performance

The final research question focused on the relationship between a preservice teachers' ability to detect effective teaching interactions (VAIL) and their ability to enact effective teaching interactions (CLASS). Following the literature on the CLASS framework, the analysis was conducted using the three class domains: emotional support, classroom organization, and instructional support. Each of these domain scores served as an outcome variable in one of three separate regressions.

Prior to examining the regression results, analysis was conducted to determine if there was a significant difference between the participants with exit VAIL sum scores that also had GRE scores and those that did not. ANOVA indicated that there was no significant difference between these two groups in either exit VAIL (F=.887, p=.348), or CLASS scores in the emotional support domain (F=.137, p=.712), classroom organization domain (F=.298, p=.586), and instructional support domain (F=1.413, p=.237).

The first predictor variable entered into each of the three regressions was the exit VAIL score. The exit VAIL score was taken from those participants who completed the VAIL in the final year of their teacher education program and had also completed their student teaching experience. These preservice teachers would have completed this

administration of the VAIL following a one-semester fall student teaching placement. For some of these individuals it may have been the first time they had taken the VAIL, while others would have taken the VAIL previously; however, each individual had completed the same amount of their respective programs. The specific individual and programmatic predictors were then entered following the exit VAIL similar to the previous two regression analyses: gender, race, academic aptitude, teaching specialty, and teacher education program. The results for all three regression equations are illustrated in Table 9.

Each regression equation produced unique results and will be discussed individually. Across the models, much of the explained variance was found in teaching specialty. Two of the total models were not significant, but one did show significant predictive quality. The emotional support regression will be discussed first followed by classroom organization and then instructional support.

Observed Teaching Quality: Emotional Support

With the CLASS emotional support as the outcome variable, the entire model was not statistically significant (adjusted R^2=.056, *p*=.204). Individual predictors were largely non-significant in the model. The Exit VAIL score did not show a significant relationship with CLASS emotional support (ß= .150, *p*=.219) in this model.

Gender, race, and performance on the GRE were not significant predictors of observed teaching performance-emotional support domain. Certain teaching specialties did demonstrate a difference in their teaching performance compared to the elementary preservice teachers. Math preservice teachers demonstrated lower scores on the emotional support domain of teaching performance than elementary teachers (ß= -.245,

Table 9
Standardized Regression Coefficients: CLASS

Total N=92	Observed Teaching Quality		
Predictors	Emotional support	Classroom Organization	Instructional support
	ß	ß	ß
Exit VAIL	.150	.035	.159
Male	.101	-.090	.273*
Ethnicity (Comparison=Caucasian)			
African American	-.009	.152	-.064
Asian	-.093	-.061	-.040
Hispanic	-.098	.029	-.075
Other/Not Specified	-.049	.139	-.090
Academic Aptitude			
GRE Verbal	.027	.183	-.011
GRE Quantitative	.148	.002	-.015
GRE Analytical Writing	-.045	.057	.104
Teaching Specialty (Comparison=Elementary)			
English	.075	.128	-.173
Foreign Language	-.084	.010	-.048
Health/PE	.049	-.017	.028
Math	-.245*	.173	-.270*
Science	-.216	.129	-.144
Special Education	.233*	.318*	.191
Social Studies	-.133	-.068	-.091
Teacher Education Program (Comparison=BAMT)			
PGMT	-.156	-.311*	-.138
Final adjusted R^2	.056	.045	.125*

*p<.05

p=.032). Whereas special education preservice teachers showed a strong statistical difference from elementary teachers. Special education preservice teachers showed higher scores (ß= .233, *p*=.055) in emotional support than elementary teachers. Differences between BAMT and PGMT students were also examined on emotional support and there were no significant differences between the two groups.

Observed Teaching Quality: Classroom Organization

In the next regression, the classroom organization domain of the CLASS was used as the outcome variable with the same predictor variables as above. Similar to the emotional support regression, the classroom organization entire model was not statistically significant (adjusted R^2=.045, *p*=.247). The Exit VAIL score was not a significant predictor of teaching performance in the classroom organization domain (ß= -.035, *p*=.774).

The specific individual characteristics did not show a significant relationship with teaching performance in the classroom organization domain. Gender, racial category, and GRE scores were all non-significant.

Conversely, programmatic characteristics of preservice teachers were significant predictors of teaching performance in the classroom organization domain. With elementary preservice teachers as the comparison group, special education (ß= .318, *p*=.010) preservice teachers were rated higher in their teaching performance.

A comparison between BAMT and PGMT teacher education students also demonstrated differences in teaching performance in the classroom organization domain. PGMT teacher education students scored significantly lower (ß= -.311, *p*=.014) than BAMT students in ratings of their teaching performance.

Observed Teaching Quality: Instructional Support

The final regression equation examined the instructional support CLASS domain as the dependent variable with the same predictors as described previously. In this instance, the entire model was statistically significant (Adjusted R^2= .125, *p*=.050). The Similar to the other domains, the exit VAIL score was not associated with teaching performance in the instructional support domain (ß= .159, *p*=.176) in this analysis. Other explained variance can be found in both the individual and programmatic characteristics.

Among the individual characteristics, gender was a significant predictor of teaching performance in the instructional support domain. In this sample, male preservice teachers were scored higher than females (ß= .273, *p*=.020). Meanwhile racial category and GRE scores had no significant relationship to teaching performance.

With elementary preservice teachers as the comparison group, some teaching specialties were scored higher and others lower. However, only math preservice teachers scored significantly different, demonstrating lower scores in instructional support than elementary teachers (ß= -.270, *p*=.015). All other teaching specialties had no significant differences than the elementary group.

In this regression equation BAMT and PGMT students were once again compared. There were no significant differences between the two groups in teaching performance on in the instructional support domain.

The analysis described in this section did find that the ability to detect effective teaching interactions combined with gender, race, academic aptitude, teaching specialty, and teacher education program were related to observed teaching performance. Analysis presented in this section also indicated that individual characteristics tended not to be

significant while teaching specialty and teacher education program had significant relationships with teaching performance. However, the teaching specialties and teacher education program characteristics that were significant varied across the domains of observed teaching interactions.

CHAPTER FIVE

DISCUSSION

This study explored the ability of a standardized instrument to measure preservice teachers' ability to detect effective teaching interactions in video-recorded segments of pre-school classes. The purpose of the study was to analyze preservice teachers' competence to detect effective teaching interactions using the Video Assessment of Interactions and Learning (VAIL). The primary questions of the research were: 1) do individual demographic characteristics (i.e. race, age, academic aptitude) or teacher education program (teaching specialty, five year bachelor's plus master's program, post-graduate program, or non-teacher education student) predict participants' ability to detect effective teaching interactions; 2) after controlling for initial performance do individual demographic or teacher education program characteristics predict performance on a final measure of participants abilities to detect effective teaching interactions; and 3) is there a relationship between participants' ability to detect effective teaching interactions and their observed teaching performance?

Through the use of descriptive statistics, correlations, and linear regression, these questions have been examined. Results paint a somewhat clouded picture. This study demonstrates the possibility of capturing the variability in preservice teacher performance

and also shows a change in VAIL scores over time; however, this variance is not well-explained by the analytic models used. There is evidence of a relationship between the ability to detect effective teaching interactions and the competence to perform those interactions as a teacher, which provides support for the use of the VAIL as a measure of teacher knowledge.

This final chapter connects the current study with prior and future research related to measuring the effects of teacher education programs on preservice teacher knowledge and skills. The first section of the chapter will summarize the major findings of the study and how these findings connect to the literature. The second section will discuss the limitations of the present study and will discuss future directions for research using the measures described in this study. Finally, the third section will discuss the implications of this research.

Summary of Findings

The analyses presented in this study suggest three main findings relating to the study's research questions. The first finding suggests that it is possible to capture the variance in preservice teachers' ability to detect effective teaching interactions in a standardized way. Second, the VAIL scores changed in those participants who took the assessment multiple times; however, these changes did not occur in predictable ways. Third, the ability to detect effective teaching interactions is associated with observed teaching performance. Taken together these findings provide preliminary support in demonstrating the capacity of a standardized measure to assess preservice teachers' ability to detect effective teaching interactions. Each of these findings will be discussed in detail this section.

Capturing Variance in Detecting Teaching Interactions

This paper argues that to identify the effects of a teacher education program on preservice teacher learning, it is important to isolate the characteristics that individuals bring with them upon program entry. This study examined participants' prior ability to detect effective teaching interactions and specific individual characteristics including gender, race, academic aptitude, teaching specialty, and teacher education program. Results indicated that a standardized measure could effectively assess preservice teachers' skills at detecting effective teaching interactions at program entry. It also showed that the specific individual characteristics studied were not related to the ability to detect effective interactions at program entry or later in the teacher education program.

Analyses conducted in this study indicate that it is possible to measure preservice teachers' ability to detect effective teaching interactions using a standardized measure. Furthermore, although not a focus of this study, this study did reinforce previous findings (Wiens et al., 2013) that the VAIL as a standardized measure can be implemented reliably across multiple time points prior to, and during, the teacher education program. This addresses an important drawback of previously used measures in teacher education which have not been used to demonstrate preservice teacher learning. Previous measures in teacher education such as student teaching evaluations (Greenberg, Pomerance, & Walsh, 2011), portfolios (Anderson & DeMuelle, 1998; Zeichner & Wray, 2001) and teacher performance assessments (Darling-Hammond, 2006; Pecheone & Chung, 2006) are end of program measures that can provide useful data on preservice teachers' preparedness to teach, but cannot provide data on learning throughout the program. However, this study indicates that using a standardized measure to investigate preservice

teacher learning is possible through the administration of the same assessment several times in a teacher education program.

While this study did find that using the same measure repeatedly can potentially help control for prior ability the individual and program characteristics were not found to be associated with competence in detecting interactions. Gender, race, academic aptitude, teaching specialty, and teacher education program did not predict participants' ability to detect effective teaching interactions. Previous research on teaching has shown that gender (Duffy, Warren, & Walsh, 2001; Good, Sikes, & Brophy, 1973) and race (Hawkins, Von Cleve, & Catalano, 1991; Zimmerman, Khoury, Vega, Gil & Warheit, 1995) can impact teacher student interactions in the classroom. This may indicate that the ability to detect effective interactions addresses a different set of skills, or a portion of the skills necessary for effective teaching. An example of this could be that measuring the skills at detecting interactions removes the socio-emotional component of the teaching exercise and therefore not all individual characteristics that may impact teaching performance would also impact the ability to detect these interactions in videos.

Likewise, there is also evidence to support that academic aptitude can be related to teaching performance (Andrew, Colbb and Giampietro, 2005); however, it was not shown to be associated with the ability to detect teaching interactions in this study. The limited variance in demonstrated academic aptitude may account for this non-significant relationship. While it may make sense that socio-emotional aspects of teaching do not translate to the measure used in this study, academic aptitude may reasonably be expected to predict this ability. Using a measure of academic aptitude, or preservice teacher

sample that contains greater variance in measured academic aptitude may find different results than presented here.

The lack of predictive power of individual and program characteristics is especially noteworthy because the measure used three videos of preschool language arts lessons. In order to understand the validity of the VAIL across teaching specialties it is important to examine if the measure is biased towards specific groups. Since the videos are of preschool language arts lessons, it is reasonable to theorize that elementary teachers or English teachers may perform better on the VAIL. However, at entry there were no differences among participants in their ability to detect effective interactions in these preschool lessons. This indicates that the VAIL does not show bias towards any particular group of participants and can be used across teaching specialties and teacher education programs. Therefore, the VAIL could potentially provide a good baseline measure of preservice teachers' competence in detecting effective interactions in place of standardized test scores and grade point averages.

Examining Change in Detecting Effective Interactions

This study sought to examine participants' change in ability to detect effective teaching interactions during their teacher education program and to understand if specific individual demographic and programmatic characteristics still were associated with a final VAIL score after controlling for the initial VAIL score. There was a large range of change between the first and last administrations of the VAIL. Additionally, there were an equal amount of participants that showed score improvement as demonstrated diminished scores. This is a concerning finding because, theoretically, preservice teachers would improve their ability to detect effective teaching interactions as they move

through the teacher education program. The analysis conducted in this study did not find that scores were higher the second time participants took the VAIL and this is something that warrants continued study in the future.

Due to the nature of the data, this analysis does not distinguish between when participants took the VAIL the first time. Likewise, when participants took it the VAIL the final time could also fluctuate—a necessity to analyze longitudinal results. The nature of the data may explain why there was not relationship between the first and second VAIL. The noise that a lack of a common initial VAIL administration provides in this analysis, will direct future researchers to hone in on this important variable and parse out the impact of the teacher education program.

Association between Detecting Interactions and Teaching Performance

Analyses presented in this study demonstrate a connection between the ability to detect effective teaching interactions and the ability to conduct effective teaching interactions. This relationship was demonstrated in significant correlations the VAIL measure taken at the end of the teacher education program and CLASS data collected during the student teaching experience. Performance on the VAIL was positively correlated with observed teaching performance in both the emotional support and instructional support domains.

These relationships did not show up in regression equations examining the predictive ability of VAIL and individual and program characteristics on the different CLASS domains. One possible explanation for this is the loss of statistical power due to lost cases in the regression analysis. Due to the list wise deletion strategy used in PASW 18, all cases without data in each variable are deleted from the analysis. Therefore, with

the addition of more variables in the regression equation, more cases are lost due to missing data. Missing GRE scores accounted for the dropped from the analysis because they did not meet the requirements of having data in each field being examined. This accounts for the drop from 130 cases in the bivariate correlation analysis to the 92 cases in the regression analysis. Analysis of variance showed that the differences between those cases dropped and those that were not dropped were not significant in their exit VAIL sum score or between any of the three CLASS domains. This may indicate that the lack of a significant relationship between the VAIL and CLASS domains in the regression analysis results from a loss of statistical power due to a reduction in available cases. The significant correlation is an important finding that shows a relationship between detecting interactions and teaching performance.

Previous research has also found a relationship between teaching performance and the ability to detect effective interactions (Hamre et al., 2012; Jamil, Sabol, et al, under review). In this study, participant's ability to detect effective teaching interactions was correlated with teaching performance in the emotional support and instructional support domains. Previously, Jamil, Sabol and colleagues (under review) found that performance on the VAIL is associated with teaching performance only in the instructional support domain. Jamil, Sabol, and colleagues (under review) point out that the prompts for the VAIL measure (see Chapter 3) instruct participants to look for instructionally oriented interactions. This should prime participants to examine the videos in ways that coincide with the interactions in that domain. However, in this sample participants demonstrated a link across two domains. Future research can help to examine this link further and examine if there is a consistent connection across samples between these two measures.

The demonstrated link between detecting and implementing effective teaching interactions shows that the VAIL is providing a signal about an important constellation of skills and knowledge related to effective teaching. A measure that provides a signal of important skills and knowledge, even when taken from a brief or narrowly focused measure, can provide important information about an individual's knowledge and skills related to larger constructs and skills related to effective teaching interactions (Gage & Needels, 1989). For example, Grossman and colleagues (2010) used the CLASS system to observe teachers six times throughout a school year and found that these six lessons were predictive of student achievement for the entire year. Grossman and colleagues (2010) followed established procedures of observing a classroom in 15 minute segments in order to rate on the CLASS scale. This demonstrates that observing short segments of lessons can provide important information about teaching interactions that indicates teaching practice beyond the observed lesson.

The VAIL builds on the empirical and theoretical support behind the CLASS framework for understanding teaching interactions (Grossman et al., 2010; Hamre & Pianta, 2007; Hamre, Pianta, Mashburn, & Downer, 2007). Performance on the VAIL has been shown to be associated with teaching performance as measured by CLASS. Then it follows that performance on the VAIL is representing important teaching knowledge and skills that matter to student learning.

The complexity of teaching (Anderson, Evertson, & Brophy; 1979; Lampert, 1985; Roth, Masciotra, & Boyd, 1999) precludes an individual measure from capturing all of the components necessary for effectiveness. However, the VAIL may provide an indication of a skill set that is present in preservice teachers that relates to teaching

behaviors that matter for student learning. This finding builds on previous research that demonstrates the connection between knowledge of teaching and teaching effectiveness (Loughram et al., 2004; Masters et al., 2010; Hill et al., 2005). As a new measure, the VAIL offers promise as a proxy for participants' ability to implement effective teaching interactions.

Measuring the impact of teacher education programs on preservice teacher learning is complicated and there is a lack of standardized measures that can be used across settings (Zeichner, 2005). The lack of authentic teaching experiences early in teacher education programs (AACTE, 2010) makes the use of standardized measures impossible for demonstrating preservice teacher learning. A standardized measure of the ability to detect effective teaching interactions that is associated with effective teaching practice can be used as a proxy for the measuring teaching performance. The VAIL is an example of standardized measure that can be implemented prior to entry in a teacher education program and again later in the program and performance on the VAIL is associated with teaching performance in a student teaching experience.

Conclusions

Teachers are important to student success in schools and represent the most important in-school factor predicting student achievement (Goldhaber, 2007; Nye et al., 2004; Rivkin, et al., 2005; Rockoff, 2004). Ensuring quality teachers are available to all students could help raise student learning and improve student achievement. The training of new teachers prior to assuming the responsibilities of teaching has fallen largely on traditional teacher education programs such as the one described in this study (National Research Council, 2010). However, teacher education programs take different

approaches to the training of preservice teachers (Boyd et al., 2007; Levine, 2006) and research has not provided strong empirical support for any specific model.

To begin to build a knowledge base of what works in teacher education, it is important to begin to develop standardized measures that can be used across contexts. This will allow for the comparison of different approaches and the effects of these teacher education models on student learning. This study demonstrates that a standardized measure can be implemented in a teacher education program across different teaching specialties and two concurrent teacher education programs. The study also shows how a standardized measure, linked to important teaching behaviors can be used to understand the knowledge of teachers at program entry and control for that knowledge to begin to potentially isolate teacher education effects. This marks an important preliminary step forward in beginning to build an empirically supported knowledge base in understanding the effects of teacher education programs on preservice teacher learning.

This study demonstrates the possibility of measuring preservice teachers' ability to detect effective teaching interactions in a standardized way. The VAIL shows promise as a standardized measure of teacher education that can be implemented reliably and capture the variance in preservice teachers' ability to detect effective teaching interactions in video-recordings. VAIL scores also changed from one administration of the measure to the next requiring further analysis to determine the nature of these changes. This study demonstrated the association between the ability of participants to detect effective teaching performance and their ability to enact effective teaching interactions in a student teaching setting indicating that the VAIL is signaling the presence of an important constellation of skills related to effective teaching. Taken together these

findings provide initial support for the ability to measure preservice teacher's ability to detect effective teaching interactions using a standardized measure.

Limitations and Future Directions

This study provides information about the effectiveness of a standardized, video-based assessment to measure participants' ability to detect effective teaching interactions. While the analysis in this study provides preliminary evidence that the VAIL can be an effective standardized measure, there are certain limitations to this study that must be considered. Future research studies on preservice teachers' ability to detect effective teaching interactions should address these limitations. In addition, specific future directions will be discussed.

This study employed the VAIL as a measure of participants' ability to detect effective teaching interactions. However, the VAIL measures a small part of the complexity of effective teaching, particularly an understanding of effective teacher student interactions and the ability to recognize these interactions in video-recorded teaching segments. There is evidence in this study and others (Hamre et al., 2012; Jamil, Sabol et al., under review) that the VAIL signals a set of knowledge and skills related to effective teacher student interactions which is particularly strong in the instructional support domain. However, there is need for further validation of the VAIL in this regard.

Examining the construction of this study, it is important to recognize that the participants in this study were not randomly sampled into the study. All of the participants self-selected into either the introductory course or one of the two teacher education programs (five-year BAMT or two-year PGMT). Moreover, the participants in this study are all students at a highly selective university. Therefore, the ability to

generalize the findings of this study is severely limited. There was no attempt to employ a random sample, nor would that be a reasonable expectation since it is impossible to randomly assign university students into a teacher education program. An alternative to random sampling of participants into teacher education programs would be to administer the VAIL in more teacher education programs. Particularly teacher education programs that contain a greater range of preservice teachers based on academic and demographic characteristics. Including more programs and participants in a study employing the VAIL would begin to build an understanding of its generalizability to more different teacher education programs and preservice teachers.

This study also lacks a control group. The analysis described focus mostly on teacher education students. When examining change in ability over time, the vast majority of the students were exposed to some portion of the intervention—in this case the teacher education program. Even when using a pretest-posttest model it is still difficult to make causal assumptions. Any change in ability may be due to the teacher education program, but it may also be the process of maturation or some other unseen factor. Including a control group in a quasi-experimental design would help to come closer to causal claims. One approach to including a control group would be to follow-up with those students who participate in the VAIL in the introductory course, but did not enroll in the teacher education program. This group would provide a reasonable control group because the introductory course does not include instruction in teaching methods. An even stronger control group would be a group of randomly selected university students who may have no interest in taking the introductory course. The introduction of a control group for longitudinal analysis would strengthen future research efforts.

An additional limitation to this study—particularly the longitudinal analysis—is the limited longitudinal nature of the data. This study examined change between two time points in the teacher education program. However, those two time points were not the same for every participant. A cleaner analysis for the BAMT students would examine scores of participants in the introductory course (pretest) and then compare those results with the same participants scores in the fourth year of the program and then again in the fifth and final year of the teacher education program. Such an analytic strategy would require four years of data to capture data on participants in their second year through fifth year of a five-year program. Unfortunately, the currently available dataset only includes three years of data. Therefore, examination of a pretest-posttest design was not possible at the time of this study. As more data is collected, future research can examine the full spectrum of data from across the teacher education program.

The narrow focus on predictive components of preservice teachers' ability to detect effective teaching interactions presents another limitation. The analysis showed that there is large variability between individuals in their competence in this area. However, a model including only gender, race, academic aptitude, teaching specialty, and teacher education program did not account for any meaningful amount of the variance in those scores. There are many skills required for understanding the importance of interpersonal interactions in effective teaching, thus, a simple model should not be able to explain much of the variation between abilities to detect effective teaching interactions. Future studies with the VAIL should begin to include a wider range of preservice teacher characteristics such as socio-emotional competences; knowledge of child development and learning; experiences working with children; and dispositions related to teaching and

learning. As more aspects of individual teacher characteristics are examined with performance on the VAIL it may be possible to explain more of the variance between individuals. This type of information could be very helpful for designing teacher education programs to meet the needs of specific preservice teachers. If individual characteristics predispose preservice teachers to understand teaching in specific ways, teacher education programs can identify this early in the program and build in scaffolds to help individuals develop these characteristics and become effective teachers or counsel individuals out of the profession.

This study does not attempt to explain why different demographic or program characteristics may be associated with the ability to detect effective teaching interactions—a final limitation. The VAIL instrument is not designed to examine the mechanisms that make an individual more or less likely to have higher or lower score. This study found that health/PE preservice teachers had lower scores on the VAIL than did elementary teachers. However, the collected data cannot begin to understand why that may be the case. Certainly the two groups take different pedagogy and subject area courses, but this study does not fully explain why one group is statistically more likely to demonstrate an ability to detect teaching interactions. Further research is required to examine differences in programs and individuals to develop a more complete understanding of why some groups may outperform other groups. Conducting direct analysis of what is being taught in different teaching specialties and how these preservice teachers are mentored could shed light on why some teachers demonstrate higher performance on these measures. Analysis of observed teaching performance reinforces

this need as some teaching specialties demonstrated more effective teaching interactions in the different CLASS domains.

Implications

This study has provided findings related to using a standardized measure in a teacher education program. The findings described in this chapter lead to three important implications for research in teacher education.

1. Standardized measurement of teacher education is possible. Previous research in teacher education has not demonstrated that standardized measurement can assess the effects of teacher education on preservice teachers. Teaching portfolios (Anderson & DeMeulle, 1998; Bannik, 2009; Wolf & Dietz, 1998) have shown the potential for this, but have not been used as a standardized measure across teacher education settings. Evaluations of teaching performance using a standardized observation measure (Wiens, under review; Jamil et al., in press) or teacher performance assessments (Darling-Hammond, 2006; Pecheone & Chung, 2006) have demonstrated that standardized measures can be applied as end of program measures.

This study shows that a video-based, standardized measure of the ability to detect interactions can be implemented in a teacher education setting and that the VAIL is able to assess the variance in preservice teachers' abilities. The VAIL represents a measure that is reasonably inexpensive, less time consuming, and can be administered several times throughout a teacher education program including as a pretest-posttest assessment. While the results of the analysis described in this study could only be considered preliminary findings, the VAIL shows promise as a standardized measure that could help

begin to show the influence of teacher education programs on the skills and knowledge of preservice teachers.

2. A Common Language is Possible in Teacher Education. Grossman and McDonald (2008) make the argument that research in teacher education has been disjointed and that it lacks a common language that would allow researchers to communicate across settings. Easton (2012) said that measurement can help to provide that common language. In education generally, and teacher education specifically, a standardized observation system like the CLASS can begin to provide that language. The CLASS (Pianta et al., 2012; Pianta et al., 2008) provides one lens on effective teaching and organizes teacher-student interactions into domains and indicators of specific teaching behaviors that make a difference to student learning. Because VAIL is based on the CLASS framework it taps into the language and organization of teaching interactions. The analysis in this study demonstrated that performance on the VAIL does predict observed teaching performance. Therefore, VAIL can contribute to a common language that can be used across settings through its consistency with the CLASS system. With a common language teacher education programs in different settings and with different designs can be compared and analyzed. This information can begin to build a knowledge base of what practices are the most effective in teacher education.

3. Multiple Measures are Necessary in Teacher Education. Teaching is complex (Anderson et al., 1979; Lampert, 1985; Roth et al., 1999). It is unreasonable to assume that a single measure could capture the complexity of teaching or learning to teach. The CLASS provides one way to measure and understand effective teaching. Analysis in this paper found that the VAIL maps onto one of the three CLASS domains. However,

neither of these instruments alone or in combination can hope to capture the complexities of learning to teach. Multiple measures are required. As the Gates Foundation (2012) showed in the Measures of Effective Teaching the most effective way to evaluate a teacher was through a variety of measures including student achievement scores, student evaluations of teachers, and several standardized observation measures. Measuring teacher education is even more complex than studying effective teaching and therefore multiple measures are certainly necessary.

CLASS and VAIL, through the examination of teacher-student interactions, do measure important constructs that impact student learning. However, they do not measure other important aspects of teaching. In addition to the socio-emotional and cognitive constructs measured by the CLASS, other important knowledge, skills and dispositions include designing and implementing curriculum effectively (Thornton, 2008), instructional preparation (Ball & Cohen, 1999; Estes et al., 2011), or continual reflection and improvement year after year (Baird et al., 1991; Schon,1987). The design and implementation of standardized measures that assess growth in these knowledge and skills through a teacher education program could also contribute to our understanding of the impacts of programs on preservice teacher learning.

Conclusion

The title of this dissertation asks, "Are they learning?" Data presented and analyzed in this study cannot provide a definitive answer to that question. This study used the ability to detect effective teaching interactions as a means for representing the learning of a complex constellation of skills related to effective teaching. Longitudinal analysis of preservice teachers' ability to detect effective teaching interactions was

inconclusive. However, this study was able to address an important problem in the research literature. This study examined whether a standardized measure could be used to measure teacher education and found sufficient evidence to suggest that the Video Assessment of Interactions in Learning could be used for this purpose. Continued study of the VAIL is necessary, but the measure shows promise in a field in need of more standardized measures.

REFERENCES

Adler, S. (2008). The education of social studies teachers. In L. Levstik & C. Tyson (Eds.), *Handbook of research in social studies education* (pp. 329-351). New York, NY: Routledge.

Ainsworth, M.D.S., Blehar, M.C., Waters, E., & Wall, S. (1978). *Patterns of attachment: A psychological study of the strange situation.* Hillside, NJ: Lawrence Erlbaum.

Allen, J.P., Pianta, R.C., Gregory, A., Mikami, A.Y., & Lun, J. (2011). An interaction-based approach to enhancing secondary school instruction and student achievement. *Science, 333,* 1034-1037.

American Association of Colleges of Teacher Education (2010). *The clinical preparation of teachers: A policy brief.* Washington, DC: Author.

American Association of Colleges of Teacher Education (2012). Selectivity of educator preparation programs. Washington, DC: Author.

Anderson, L.M., Evertson, C.M., & Brophy, J.E. (1979). An experimental study of effective teaching in first-grade reading groups. *The Elementary School Journal, 79*(4), 193-223.

Anderson, R.S. & DeMeulle, L. (1998). Portfolio use in twenty-four teacher education programs. *Teacher Education Quarterly, 25,* 23-31.

Andrew, M.D., Cobb, C.D., & Giampietro, P.J. (2005). Verbal ability and teacher effectiveness. *Journal of Teacher Education, 56,* 343-354.

Arnold, D.H., McWilliams, L. & Arnold, E.H. (1998). Teacher discipline and child misbehavior in day care: Untangling causality with correlational data. *Developmental Psychology, 34,*276-287.

Association of Teacher Educators (2000). *Standards for field experiences in teacher education.* Reston, VA: Author.

Assor, A., Kaplan, H., & Roth, G. (2002). Choice is good, but relevance is excellent: Autonomy-enhancing and suppressing teacher behaviors predicting students' engagement in schoolwork. *British Journal of Educational Psychology, 72,* 261-278.

Attebery, A., Loeb, S., & Wyckoff, J. (2012, November) *Returns to experience in early career years.* Paper presented at the Association for Public Policy Analysis and Management Fall Conference, Baltimore, MD.

Au, W. (2007) High-stakes testing and curricular control: A qualitative metasynthesis. *Educational Researcher,* 36, 259-267.

Baird, J.R., Fensham, P.J., Gunstone, R.F., & White, R.T. (1991). The importance of reflection in improving science teaching and learning. *Journal of Research in Science Teaching, 28,* 163-182.

Baker, J.A. (1999). Teacher-student interaction in urban at-risk classrooms: Differential behavior, relationship quality, and student satisfaction with school. *The Elementary School Journal, 100*(1), 57-70.

Baker, E.L., Barton, P.E., Darling-Hammond, L., Haertel, E., Ladd, H.F., Linn, R.L., … Sheppard, L.A. (2010). *Problems with the use of student test scores to evaluate teachers* (EPI Briefing Paper 278). Washington, DC: Economic Policy Institute.

Baker, C.N., Kupersmidt, J.B., Voegler-Lee, M.E., Arnold, D.H., Willoughby, M.T. (2010). Predicting teacher participation in a classroom-based, integrated preventive intervention for preschoolers. *Early Childhood Research Quarterly, 25,* 270-283.

Ball, D.L. & Cohen, D.K. (1999). Developing practice, developing practitioners: Toward a practice-based theory of professional education. In L. Darling-Hammond & G. Sykes (Eds.) *Teaching as the learning profession: Handbook of policy and practice.* (pp.3-32). San Francisco, CA: Jossey-Bass Publishers.

Ballou, D., Sanders, W., & Wright, P. (2004). Controlling for student background in value-added assessment of teachers. *Journal of Educational and Behavioral Statistics, 29*(1), 37-65.

Bannink, A. (2009). How to capture growth? – Video narratives as an instrument in teacher education. *Teaching and Teacher Education, 25,* 244-250.

Bennett, S.K. (1982). Student perceptions of and expectations for male and female instructors: Evidence relating to the question of gender bias in teacher evaluation. *Journal of Educational Psychology, 74,* 170-179.

Berliner, D.C. (1979). Tempus educare. In Peterson, P.L. & Walberg, H.J. (Eds.) *Research on Teaching* (pp 120-135). Berkeley, CA: McCutchan.

Berrill, D.P. & Addison, E. (2010). Repertoires of practice: Re-framing teaching portfolios. *Teaching and Teacher Education, 26,* 1178-1185.

Borman, G.D., & Kimball, S.M. (2005). Teacher quality and educational equity: Do teachers with higher standards-based evaluation ratings close the achievement gaps? *The Elementary School Journal*, *106*(1), 3-20.

Boyd, D., Grossman, P., Lankford, H., Loeb, S., & Wyckoff, J. (2005). *How changes in entry requirements alter the teacher workforce and affect student achievement*. (NBER Working Paper No. 11844). Cambridge, MA: National Bureau of Economic Research.

Boyd, D., Goldhaber, D., Lankford, H., & Wyckoff, J. (2007). The effect of certification and preparation on teacher quality. *The Future of Children 17*(1): 45-68.

Boyd, D., Lankford, H., Loeb, S., Rockoff, J., & Wyckoff, J. (2008). The narrowing gap in New York City teacher qualifications and its implications for student achievement in high-poverty schools. *Journal of Policy Analysis and Management, 27,* 793-818.

Bransford, J., Darling-Hammond, L., & LePage, P. (2005). Introduction. In Darling-Hammond, L., & Bransford, J. (Eds.) *Preparing teachers for a changing world: What teachers should learn and be able to do* (pp 1-39). San Francisco, CA: Jossey-Bass.1-39.

Braun, H.I. (2005). *Using student progress to evaluate teachers: A primer on value-added models.* Princeton, NJ: Educational Testing Service.

Bretherton, I. (1985). Attachment theory: Retrospect and prospect. *Monographs of the Society for Research in Child Development, 50*(1/2), 3-35.

Bronfenbrenner, U. (1993). Ecological models of human development. In Gauvain, M. & Cole, M. (Eds.), *Readings on the development of children, 2nd Ed.* (pp. 37-43). New York, NY: Freeman.

Brophy, J. E., & Good, T. L. (1986). Teacher behavior and student achievement. In M. C. Wittrock (Ed.), *Handbook of research on teaching* (3rd edition, pp. 328-375). New York, NY: Macmillan.

Brophy, J. E. (1999). Teaching (Educational Practices Series-1). Geneva, Switzerland: International Academy of Education and International Bureau of Education, UNESCO.

Burchinal, M., Roberts, J.E., Zeisel, S.A., Hennon, E.A., & Hooper, S. (2006). Social risk and protective child, parenting, and child care factors in early elementary school years. *Parenting: Science and Practice, 6*(1), 79-113.

Butin, D. (2005). How social foundations of education matters to teacher preparation: A policy brief. *Educational Studies, 38*(3), 214-229.

Cadima, J., Leal, T, & Burchinal, M. (2010). The quality of teacher-student interactions: Associations with first graders' academic and behavioral outcomes. *Journal of School Psychology, 48,* 457-482.

Carroll, D. (2012). Examining the development of dispositions for ambitious teaching: One teacher candidate's journey. *The New Educator, 8*(1), 38-64.

Chetty, R., Friedman, J.N., Hilger, N., Saez, E., Schanzenbach, D.W., & Yagan, D. (2010). *How does your kindergarten classroom affect your earnings? Evidence form Project Star.* (Working Paper 16381). Cambridge, MA: National Bureau of Economic Research.

Chung, R. (2008). Beyond assessment: Performance assessments in teacher education. *Teacher Education Quarterly, 35*(1), 7-29.

Clotfelter, C.T., Ladd, H., & Vigdor, J.L. (2006). Teacher-student matching and the assessment of teacher effectiveness. *Journal of Human Resources, 41*(4), 778-820.

Clotfelter, C.T., Ladd, H., & Vigdor, J.L. (2007). *How and why do teacher credentials matter for student achievement?* Washington, DC: National Center for Analysis of Longitudinal Data in Education Research.

Cochran-Smith, M. & Lytle, S.L. (2006). Troubling images of teaching in No Child Left Behind. *Harvard Educational Review, 76,* 668-697.

Cochran-Smith, M. & Zeichner, K.M. (2005). Executive summary. In M. Cochran-Smith, & K.M. Zeichner, (Eds.). *Studying teacher education: The report of the AERA Panel on Research and Teacher Education.* (pp. 1-36). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Coleman, J.S. (1966). *Equality of educational opportunity.* Washington, DC: U.S. Department of Health, Education, and Welfare, Office of Education.

Constantine, J., Player, D., Silva, T., Hallgren, K, Grider, M., & Deke, J. (2009). *An evaluation of teachers trained through different routes to certification, final report* (NCEE 2009-4043). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Corbett, C., Hill, C., & St. Rose, A. (2008). *Where the girls are: The facts about gender equity in education.* Washington, DC: AAUW Educational Foundation.

Council of Chief State School Officers (2011). *Interstate Teacher Assessment and Support Consortium (InTASC) Model core teaching standards: A resource for state dialogue.* Washington, DC: Author.

Creswell, J.W. (2008). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research*. Upper Saddle River, NJ: Pearson.

Curby, T.W., Rimm-Kaufman, S.E., & Ponitz, C.C. (2009). Teacher-child interactions and children's achievement trajectories across kindergarten and first grade. *Journal of Educational Psychology, 101*(4), 912-925.

Danielson, C. (1996). *Enhancing professional practice: A framework for teaching.* Alexandria, VA: Association for Supervision and Curriculum Development.

Danielson, C., & McGreal, T.L. (2000). *Teacher evaluation to enhance professional practice.* Alexandria, VA: Association for Supervision and Curriculum Development.

Darling-Hammond, L. (2006). Assessing teacher education: The usefulness of multiple measures for assessing program outcomes. *Journal of Teacher Education, 57,* 120-138.

Darling-Hammond, L., Chung, R., & Frelow, F. (2002). Teacher preparation: How well do different pathways prepare teachers to teach? *Journal of Teacher Education, 53*(4), 286-302.

Darling-Hammond, L., & Bransford, J. (2005). *Preparing teachers for a changing world: What teachers should learn and be able to do.* San Francisco, CA: Jossey-Bass.

Darling-Hammond, L., Holtzman, D.J., Gatlin, S.J., & Heilig, J.V. (2005). *Does teacher preparation matter? Evidence about teacher certification, Teach for America, and teacher effectiveness.* Stanford, CA: Stanford University.

Davis, E. (2010). *Trust, triage, and teaching: Exploring interactions between mentors and first-year teachers.* (unpublished doctoral dissertation). University of Virginia, Charlottesville, VA.

Deci, E.L., & Ryan, R.M. (1987). The support of autonomy and the control of behavior. *Journal of Personality and Social Psychology, 53,* 1024-1037.

Decker, D.M., Dona, D.P., & Christenson, S.L. (2007). Behaviorally at-risk African American students: The importance of student-teacher relationships for student outcomes. *Journal of School Psychology, 45,* 83-109.

Decker, L.E., & Rimm-Kaufman, S.E. (2008). Personality characteristics and teacher beliefs among pre-service teachers. *Teacher Education Quarterly, 35*(2), 45-64.

Dedeoglu, H. & Lamme, L.L. (2011). Selected demographics, attitudes, and beliefs about diversity of preservice teachers. *Education and Urban Society, 43,* 468-485.

Douglas, K. (2009). Sharpening our focus in measuring classroom instruction. *Educational Researcher, 38*, 518-521.

Dominguez, X., Vitiello, V.E., Maier, M.F., & Greenfield, D.B. (2010). A longitudinal examination of young children's learning behavior: Child-level and classroom-level predictors of change throughout the preschool year. *School Psychology Review, 39,* 29-47.

Duffy, J. Warren, K., & Walsh, M. (2001). Classroom interactions: Gender of teacher, gender of student, and classroom subject. *Sex Roles, 45,* 579-593.

Dunkin, M. & Biddle, B. (1974). *The study of teaching.* New York, NY: Holt, Rinehart, & Winston.

Easton, J. (2012). Inaugural opening plenary session. National Council on Measurement in Education. April 14, 2012, Washington, DC.

edTPA. (2012). About edTPA. Retrieved from http://edtpa.aacte.org/about-edtpa.

Emmer, E.T., & Stough, L.M. (2001). Classroom management: A critical part of educational psychology, with implications for teacher education. *Educational Psychologist, 36*(2), 103-112.

Erickson, F. (1986). Qualitative methods in research on teaching. In M.C. Wittrock, (Ed.), *Handbook of research on teaching* (3rd edition, pp. 119-161). New York, NY: Macmillan.

Estes, T.H., Mintz, S.L., & Gunter, M.A. (2011). *Instruction: A models approach.* (6th edition). Boston, MA: Pearson.

Ewing, T. (2008). ETS-led team to investigate classroom observational tool for student-teacher interactions. Retrieved May 27, 2011 from http://www.ets.org/newsroom/news_releases/investigating_classroom_observation_tools

Feiman-Nemser, S. (2008). Teacher learning: How do teachers learn to teach? In M.

Cochran-Smith, S. Feiman-Nemser, D.J. McIntyre, & K.E. Demers, (Eds.) *Handbook of research on teacher education* (3rd Edition, pp. 697-705). New York, NY: Routledge, Taylor & Francis Group.

Flanders, N. (1970). *Analyzing teaching behavior.* Reading, MA: Addison-Wesley.

Friere, P. (2009). *Pedagogy of the oppressed.* (D. Macedo, Trans.). New York, NY: Continuum International Publishing Group, Inc.

Gage, N.L. & Needels, M.C. (1989). Process-product research on teaching: A review of criticisms. *The Elementary School Journal, 89*(3), 253-300.

Gansle, K.A., Burns, J.M., Noell, G. (2010). Value added assessment of teacher preparation in Louisiana: 2005-2006 to 2008-2009. Overview of performance bands. *Louisiana Board of Regents.* August 26, 2010.

Gansle, K.A., Noell, G.H., & Burns, J.M. (2012). Do student achievement outcomes differ across teacher preparation programs? An analysis of teacher education in Louisiana. *Journal of Teacher Education, 63,* 304-317.

Garet, M.S., Porter, A.C., Desimone, L., Birman, B.F., & Yoon, K.S. (2001). What makes professional development effective? Results from a national sample of teachers. *American Educational Research Journal, 38,* 915-945.

Garmon, M.A. (2004). Changing preservice teachers' attitudes/beliefs about diversity: What are the critical factors? *Journal of Teacher Education, 55,* 201-213.

Gates Foundation (2010). Measures of Effective Teaching Project, *Working with Teachers to Develop Fair and Reliable Measures of Effective Teaching*. Retrieved from: http://www.gatesfoundation.org/learning/Pages/2010-reliable-measures-effective-teaching.aspx

Gates Foundation (2012). *Gathering feedback for teaching: Combining high-quality observations with student surveys and achievement gains.* Seattle, WA: Author.

Gersten, R., Dimino, J., Jayanthi, M., Kim, J.S., & Santoro, L.E. (2010). Teacher study group: Impact of the professional development model on reading instruction and student outcomes in first grade classrooms. *American Educational Research Journal, 47,* 694-739.

Glazerman, S., Loeb, S., Goldhaber, D., Staiger, D., Raudenbush, S., & Whitehurst, G. (2010). *Evaluating teachers: The important role of value-added*. Washington, DC: Brown Center on Education Policy at Brookings.

Goldhaber, D. (2007). Everyone's doing it, but what does teacher testing tell us about teacher effectiveness? *Journal of Human Resources, 42*(4), 765-794.

Goldhaber, D. (2008). Teachers matter; effective teacher quality policies are elusive. In H.F. Ladd & E.B. Fiske (Eds.) *Handbook of research in education finance and policy.* (pp. 146-165) New York: Routledge.

Goldhaber, D, & Liddle, S., (2011). *The gateway to the profession: Assessing teacher preparation programs based on student achievement.* Seattle, WA: Center for Education Data and Research.

Good, T.L., Sikes, J.N., & Brophy, J.E. (1973). Effects of teacher sex and student sex on classroom interaction. *Journal of Educational Psychology, 64,* 74-87.

Goodland, J.I. (1990). *Teachers for our nation's schools.* San Francisco, CA: Jossey-Bass.

Graue, E., Raushcer, E., & Sherfinski, M. (2009). The synergy of class size reduction and classroom quality. *The Elementary School Journal, 110*(2), 178-201.

Greenberg, J., Pomerance, L., & Walsh, K. (2011). *Student teaching in the United States.* downloaded from www.nctq.org/edschoolreports/studentteaching. Washington, DC: National Council on Teacher Quality.

Grossman, P. (1990). *The making of a teacher: Teacher knowledge and teacher education.* New York: Teacher College Press.

Grossman, P. (2010). *Policy brief. Learning to practice: The design of clinical experience in teacher preparation.* Washington, DC: American Association of Colleges of Teacher Education.

Grossman, P. & McDonald, M. (2008). Back to the future: Directions for research in teaching and teacher education. *American Educational Research Journal.* 45: 184-205.

Grossman, P., Boyd, D., Brown, M., Cohen, J., Lankford, H., Loeb, S., ... & Wyckoff, J. (2010). *Measure for Measure: A Pilot Study Linking English Language Arts Instruction and Teachers' Value-Added to Student Achievement*. (Working Paper 16015). Cambridge, MA: National Bureau of Economic Research.

Guaglianone, C.L., Payne, M., Kinsey, G.W., & Chiero, R. (2009). Teaching performance assessment: A comparative study of implementation and impact amongst California State University campuses. *Issues in Teacher Education, 18*(1), 129-148.

Guba E.G. & Lincoln, Y.S. (1985). Competing paradigms in qualitative research. In N.K. Denzin & Y.S. Lincoln (Eds.), *Handbook of qualitative research* (pp. 105-117). Thousand Oaks, CA: Sage.

Hallinger, P. & Murphy, J.F. (1986). The social context of effective schools. *American Journal of Education, 94,* 328-355.

Hanushek, E.A. (2002). *The long run importance of school quality* (NBER Working Papers 9071). Cambridge, MA: National Bureau of Economic Research.

Hamre, B.K. & Pianta, R.C. (2007). Learning opportunities in pre-school and early elementary classrooms. In R. Pianta, M. Cox, & K. Snow (Eds.), *School readiness and the transition to kindergarten in the era of accountability* (49-84). Baltimore: Brookes.

Hamre, B. K., Pianta, R. C., Mashburn, A. J., & Downer, J. T. (2007). Building a science of classrooms: Application of the CLASS framework in over 4,000 U.S. early childhood and elementary classrooms. New York: Foundation for Child Development. Available at http://www.fcd-us. org/resources/resources_show.htm?doc_id=507559

Hamre, B.K., Justice, L.M., Pianta, R.C., Kilday, C., Sweeney, B., Downer, J.T., Leach, A. (2010). Implementation fidelity of MyTeachingPartner literacy and language activities: Association with preschoolers' language and literacy growth. *Early Childhood Research Quarterly, 25,* 329-347.

Hamre, B. K., Pianta, R. C., Burchinal, M., Field, S., LoCasale-Crouch, J., Downer, J. T., … Scott-Little, C. (2012). A course on effective teacher-child interactions: Effects on teacher beliefs, knowledge, and observed practice. *American Educational Research Journal, 49*(1), 88-123.

Hamre, B., Downer, J., Jamil, F., & Pianta, R. (in press) Enhancing teacher intentionality: A conceptual model for the development and testing of professional development interventions. In R. Pianta, L. Justice, S. Barnett, & S. Sheridan, (Eds.), *Handbook of early education*. New York: Guilford Publications.

Hanushek, E.A. (1992). The trade-off between child quantity and quality. *Journal of Political Economy, 100*(1), 84-117.

Hanushek, E.A., Kain, J.F., O'Brien, D.M., & Rivkin, S.G. (2005). *The market for teacherquality* (Working Paper No. 11154). Cambridge, MA: National Bureau of Economic Research.

Hartford, J. & MacRuairc, G. (2008). Engaging student teachers in meaningful reflective practice. *Teaching and Teacher Education*, 24: 1884-1892.

Hawkins, J.D., von Cleve, E., & Catalano, R.F., Jr., (1991). Reducing early childhood aggression: Results of a primary prevention program. *Journal of the American Academy of Child & Adolescent Psychiatry, 30,* 208-217.

Hill, H.C., Rowan, B., & Ball, D.L. (2005). Effects of teachers' mathematical knowledge for teaching on student achievement. *American Educational Research Journal, 42*, 371-406.

Horowitz, F.D., Darling-Hammond, L., Bransford, J., Comer, J., Roesbrock, K. Austin, K., & Rust, F. (2005). Educating teachers for developmentally appropriate practice. In L. Darling-Hammond & J. Bransford, (Eds.). *Preparing teachers for a changing world: What teachers should learn and be able to do* (pp 88-125). San Francisco, CA: Jossey-Bass.

Hughes, J.N., Cavell, T.A., & Willson, V. (2001). Further support for the developmental significance of the quality of the teacher-student relationship. *Journal of School Psychology, 39*(4), 389-301.

Humphrey, D.C. & Wechsler, M.E. (2007). Insights into alternative certification: Initial findings from a national study. *Teachers College Record, 109,* 483-530.

Ingersoll, R.M. & Smith, T.M. (2004). What are the effects of induction and mentoring on beginning teacher turnover? *American Educational Research Journal, 41*(3), 681-714.

Jacob, B.A. & Lefgren, L. (2005). *Principals as agents: Subjective performance measurement in education*. (NBER Working Paper no. 11463). Cambridge, MA: National Bureau of Economic Research.

Jamil, F.M., Sabol, T.J., Hamre, B.K., & Pianta, R.C. (under review). Assessing teachers' skills in detecting and identifying effective interactions in the classroom: Theory and measurement.

Jamil, F., Downer, J.T., Pianta, R.C. (in press). Association of pre-service teachers' performance, personality, and beliefs with teacher self-efficacy at program completion. *Teacher Education Quarterly.*

Jones, M.G., Jones, B.D., Hardin, B. Chapman, L., Yarbrough, T.,& Davis, M. (1999) The impact of high-stakes testing on teachers and students in North Carolina. *The Phi Delta Kappan, 81*(3), 199-203.

Kagan, D.M. (1992). Professional growth among pre-service and beginning teachers. *Review of Educational Research, 62*(2), 129-169.

Kane, T.J., Rockoff, J.E., & Staiger, D.O. (2006). *What does certification tell us about teacher effectiveness? Evidence from New York City.* (Working Paper 12155). Cambridge, MA: National Bureau of Economic Research.

Kee, A.N. (2012). Feelings of preparedness among alternatively certified teachers: What is the role of program features? *Journal of Teacher Education, 63*(1), 23-38.

Kimball, S.M., White, B., & Milanowski, A.T. (2004). Examining the relationship between teacher evaluation and student assessment results in Washoe County. *Peabody Journal of Education, 79*(4), 54-78.

Kingston, P.W., Hubbard, R., Lapp, B., Schroeder, P., & Wilson, J. (2003). Why education matters. *Sociology of Education, 76*(1), 53-70.

Klem, A.M., & Connell, J.P. (2004). Relationships matter: Linking teacher support to student engagement and achievement. *Journal of School Health, 74*(7), 262-273.

Konold, T., Jablonski, B., Nottingham, A., Kessler, L., Byrd, S., Imig, S., Berry, R., & McNergney, R. (2008). Adding value to public schools: Investigating teacher education, teaching, and pupil learning. *Journal of Teacher Education 59*(4), 300-312.

Krauss, S., Brunner, M., Kunter, M., Baumert, J., Blum, W., Neubrand, M., & Jordan, A. (2008). Pedagogical content knowledge and content knowledge of secondary mathematics teachers. *Journal of Educational Psychology, 100,* 716-725.

Kupermintz, H. (2003). Teacher effects and teacher effectiveness: A validity investigation of the Tennessee Value Added Assessment System. *Educational Evaluation and Policy Analysis, 25*(3), 287-298.

La Paro, K. M., Pianta, R. C., Stuhlman, M. (2004). The Classroom Assessment Scoring System: Findings from the prekindergarten year. *The Elementary School Journal, 104*(5), 409-426.

La Paro, K., Hamre, B., Locasale-Crouch, J., Pianta, R., Bryant, D., Early, D., Clifford, R. et al. (2009). Quality in kindergarten classrooms: Observational evidence for the need to increase children's learning opportunities in early education classrooms. *Early Education & Development, 20*(4), 657-692.

Labaree, D.F. (2008) An uneasy relationship: The history of teacher education in the university. In M. Cochran-Smith, S. Feiman-Nemser, D.J. MacIntyre, and K.E. Demers (Eds.) *Handbook of Research on Teacher Education.* (3rd Edition, pp. 290-306). New York: Routledge.

Lampert, M. (1985). How do teachers manage to teach? Perspectives on problems in practice. *Harvard Educational Review, 55,* 178-194.

Landis, J.R., & Koch, G.G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*(1), 159–174.

Lee, K., & Choi, I. (2008). Learning classroom management through web-based case instruction: Implications for early childhood teacher education. *Early Childhood Education Journal, 35,* 495-503. doi:10.1007/s10643-008-0250-7

Levine, A. (2006). *Educating school teachers.* Washington, DC: The Education Schools Project. Retrieved from www.edschools.org.

Loeb, S., & Reininger, M. (2004). *Public policy and teacher labor markets: What we know and why it matters*. East Lansing, MI: The Education Policy Center at Michigan State University.

LoCasale-Crouch, J., Konold, T., Pianta, R., Howes, C., Burchinal, M., Bryant, D., Clifford, R., Early, D., & Barbarin, O. (2007). Observed classroom quality profiles in state-funded pre-kindergarten programs and associations with teacher, program, and classroom characteristics. *Early Childhood Research Quarterly, 22*(1), 3-17.

LoGerfo, L., Nichols, A., & Chaplin, D. (2006). *Gender gaps in math and reading gains during elementary and high school by race and ethnicity.* Washington, DC: Urban Institute.

Loughran, J.J. (2002). Effective reflective practice: In search of meaning in learning about teaching. *Journal of Teacher Education,* 53: 33-43.

Loughran, J. & Corrigan, D. (1995). Teaching portfolios: A strategy for developing learning and teaching in preservice education. *Teaching and Teacher Education, 11,* 565-577.

Loughran, J., Mulhall, P., & Berry, A. (2004). In search of pedagogical content knowledge in science: Developing ways of articulating and documenting professional practice. *Journal of Research in Science Teaching, 41*(4), 370-391.

Lortie, D.C. (1975). *Schoolteacher: A sociological study.* Chicago, Il: University of Chicago Press.

Ludwig, M., Kirshstein, R., Sidana, A., Ardila-Rey, A., & Bae, Y. (2010). An emerging picture of the teacher preparation pipeline. Washington, DC: American Institutes for Research and American Association of Colleges of Teacher Education.

Macmillan, C.J.B. & Garrison, J.W. (1984). Using the "new philosophy of science" in criticizing current research traditions in education. *Educational Researcher, 13*(10), 15-21.

Malmberg, L.E., & Hagger, H. (2009). Changes in student teachers' agency beliefs during a teacher education year, and relationships with observed classroom quality, and day-to-day experiences. *British Journal of Educational Psychology,*

79(4), 677-694.

Masters, J., de Kramer, R.M., O'Dwyer, L.M., Dash, S., & Russell, M. (2010). *Journal of Educational Computing Research, 43*(3), 355-375.

Medley, D.M. & Crook, P.R. (1980). Research in teacher competency and teaching tasks. *Theory into Practice*, 29(4), 294-301.

Medley, D. & Mitzel, H. (1963). Measuring classroom behavior by systematic observation. In N.L. Gage, (Ed.), *Handbook for Research on Teaching* (pp. 247-328). Chicago: Rand McNally.

McNergney, R.F. & Imig, S.R. (2006, October 26). Challenges of assessing valued added to PK-12 education by teacher education. *Teachers College Record* [On-line]. Downloaded January 8, 2009 from http://www.tcrecord.org/PrintContent.asp?ContentID=12814.

Milanowski, A. (2004). The relationship between teacher performance evaluation scores and student achievement: Evidence from Cincinnati. *Peabody Journal of Education, 79*(4), 33-53.

Mueller, C.M. (2012). The impact of teacher certification programs on teacher efficacy, job satisfaction, and teacher performance: A comparison of traditional and alternative certification. *Dissertation.* Paper 28. http://digitalcommons.wku.edu/diss/28.

Mulligan, C.B. (1999). Galton versus the human capital approach to inheritance. *Journal of Political Economy, 107,* S184-S224.

Murnane, R.J., Willett, J.B., Duhaldeborde, Y., & Tyler, J.H. (2000). How important are the cognitive skills of teenagers in predicting subsequent earnings? *Journal of Policy Analysis and Management, 19,* 547-568.

Murray, F.B. (2008). Teaching by second nature. In M. Cochran-Smith, S. Feiman-Nemser, D.J. MacIntyre, and K.E. Demers (Eds.) *Handbook of Research on Teacher Education*. (3rd Edition, pp 1228-1246). New York, NY: Routledge.

National Commission on Teaching and America's Future. (2003). *No dream denied: A pledge to America's children.* Washington, DC: Author.

National Council on Teacher Quality (2012). *State of the states 2012: Teacher effectiveness policies.* Washington, DC: Author.

National Council for the Accreditation of Teacher Education [NCATE]. (2008). *Professional standards for the accreditation of teacher preparation institutions.* Washington, DC: Author.

National Research Council (2000). *How people learn: Brain, mind, experience, and school*. Washington, DC: National Academy Press.

National Research Council (2010). *Preparing teachers: Building evidence for sound policy*. Washington, DC: National Academy Press.

Nye, B., Konstantopoulos, S., & Hedges, L. (2004). How large are teacher effects? *Educational Evaluation and Policy Analysis, 26*, 237-257.

O'Connor, E. & McCartney, K. (2007). Examining teacher-child relationships and achievement as part of an ecological model of development. *American Educational Research Journal, 44*, 340-369.

Odden, A. (2004). Lessons learned about standards-based teacher evaluation systems. *Peabody Journal of Education, 79*(4), 126-137.

O'Donnell, C.L., (2008). Defining conceptualizing, and measuring fidelity of implementation and its relationship to outcomes in K-12 curriculum intervention research. *Review of Educational Research, 78*, 33-84.

Okhremtchouk, I., Seiki, S., Gilliland, B., Ateh, C., Wallace, M., Kato, A. (2009). Voices of pre-service teachers: Perspectives on the Performance Assessment for California Teachers (PACT), *Issues in Teacher Education, 18*(1), 39-62.

Parkarinen, E., Kiuru, N., Lerkkanen, M.-K., Poikkeus, A.-M., Siekkinen, M., & Numi, J.-E. (2010). Classroom organization and teacher stress predict learning motivation in kindergarten children. *European Journal of Psychology of Education, 25*(3), 281-300.

Pecheone, R.L. & Chung, R.R. (2006). Evidence in teacher education: The Performance Assessment for California Teachers (PACT). *Journal of Teacher Education, 57*, 22-36.

Penuel, W.R., Fishman, B.J., Yamaguchi, R., & Gallagher, L.P. (2007). What makes professional development effective? Strategies that foster curriculum implementation. *American Educational Research Journal, 44*, 921-958.

Peterson, K. (2004). Research on school teacher evaluation. *NASSP Bulletin, 88*(639), 60-79.

Pianta, R.C. (2003). *Standardized observations from pre-K to 3rd grade: A mechanism for improving access to high-quality classroom experiences and practices during the p-3 years* (Working paper). New York: Foundation for Child Development.

Pianta, R.C., Belsky, J., Houts, R., & Morrison, F. (2007). *Science, 315*, 1795-1796.

Pianta, R.C., Belsky, J., Vandergrift, N., Houts, R., & Morrison, F.J. (2008). Classroom effects on children's achievement trajectories in elementary school. *American Educational Research Journal, 45*(2), 365-397.

Pianta, R. C., & Hamre, B. K. (2009a). Conceptualization, measurement, and improvement of classroom processes: Standardized observation can leverage capacity. *Educational Researcher*, *38*(2), pp. 109-119. doi: 10.3102/0013189X09332374

Pianta, R.C. & Hamre, B.K. (2009b). A lot of students and their teachers need support: Using a common framework to observe teacher practices might help. *Educational Researcher, 38,* 546-548. doi: 10.3102/0013189X09348786

Pianta, R.C., Hamre, B.K. & Mintz, S. (2012) *Classroom Assessment Scoring System: Secondary Manual*. Charlottesville, VA: Teachstone.

Pianta, R.C., La Paro, & Hamre, B.K. (2008). *Classroom Assessment Scoring System (CLASS).* Baltimore, MD: Paul H. Brookes.

Plecki, M.L., Elfers, A.M., & Nakamura, Y. (2012). Using evidence for teacher education program improvement and accountability: An illustrative case of the role of value-added measures. *Journal of Teacher Education, 63*(5), 318-334.

Putnam, R.T. & Borko, H. (2000). What do new views of knowledge and thinking have to say about research on teacher learning? *Educational Researcher, 29*(1), 4-15.

Porter, A.C., Youngs, P., & Odden, A. (2001). Advances in teacher assessments and their uses. In V. Richardson, (Ed). *Handbook of research on teaching* (4th Edition, pp 259-297). Washington, DC: American Educational Research Association.

Rampey, B.D., Dion, G.S., & Donahue, P.L. (2009). *NAEP 2008 trends in academic progress* (NCES 2009-479). Washington, DC: National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education.

Raver, C.C. (2004). Placing emotional self-regulation in sociocultural and socioeconomic contexts. *Child Development, 75,* 346-363.

Reardon, S.F. & Galindo, C. (2009). The Hispanic-white achievement gap in math and reading in the elementary grades. *American Educational Research Journal, 46,* 853-891.

Ripski, M.B., LoCasale-Crouch, J., & Decker, L. (2011). Pre-service teachers: Dispositional traits, emotional states, and quality of teacher-student interactions. *Teacher Education Quarterly,* Spring, 77-96.

Rivkin, S., Hanushek, E.A., & Kain, J.F. (2005). Teachers, schools and academic achievement. *Econometrica, 73*(2), 417-458.

Rockoff, J.E. (2004). The impact of individual teachers on student achievement: Evidence from panel data. *American Economic Review, 94*(2), 247-252.

Roeser, R.W., Eccles, J.S., & Sameroff, A.J. (2000). School as a context of early adolescents' academic and social-emotional development: A summary of research findings. *The Elementary School Journal, 100,* 443-471.

Rogoff, B. (2003). *The cultural nature of human development.* New York, NY: Oxford University Press.

Roth, W., Masciotra, D., & Boyd, N. (1999). Becoming-in-the-classroom: A case study of teacher development through coteaching. *Teaching and Teacher Education, 15,* 771-784.

Ryan, R.M. & Deci, E.L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist, 55*(1), 68-78.

Sanders, W.L. (2000). Value-added assessment from student achievement data: Opportunities and hurdles. *Journal of Personnel Evaluation in Education, 14*(4), 329-339.

Sanders, W., & Horn, S. (1998). Research findings from the Tennessee Value-added Assessment System (TVAAS) database: Implications for educational evaluation and research. *Journal of Personnel Evaluation in Education, 13*(3), 247-256.

Schon, D. (1983). *The Reflective Practitioner.* New York: Basic Books.

Schon, D. (1987). *Educating the Reflective Practitioner.* San Francisco: Jossey-Bass.

Schulman, L.S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher, 15*(2), 4-14.

Schulman, L.S. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard Educational Review, 57,* 1-22.

Schwartz, D.L. & Bransford, J.D. (1998). A time for telling. *Cognition and Instruction, 16,* 475-522.

Silva, D.Y., & Mason, T.C. (2003). Developing pedagogical content knowledge for civics in elementary teacher education. *Theory and Research in Social Education, 31,* 366-397.

Stefanou, C.R., Perencevich, K.C., DiCintio, M., & Turner, J.C. (2004). Supporting autonomy in the classroom: Ways teachers encourage student decision making and ownership. *Educational Psychologist, 39*(2), 97-110.

Stone, C.A. & Lane, S. (2003). Consequences of a state accountability program: Examining relationships between school performance gains and teacher, student, and school variables. *Applied Measurement in Education, 16*(1), 1-26.

Swain, S.S. (1998). Studying teachers' transformations: Reflection as methodology. *Clearing House. 72*(1), 28-34.

Taylor, B.M., Pearson, P.D., Peterson, D.S., & Rodriguez, M.C. (2003). Reading growth in high-poverty classrooms: The influence of teacher practices that encourage cognitive engagement in literacy learning. *The Elementary School Journal, 104*, 3-28.

Thornton, S.J. (2008). Continuity and change in social studies curriculum. In L.S. Levstik & C.A. Tyson (Eds.) *Handbook of research in social studies education.* (pp. 15-32). New York: Routledge.

Tigelaar, D.E.H., Dolmans, D.H.J.M., Wolfhagen, I.H.A.P, & van der Vleuten, C.P.M. (2004). Using a conceptual framework and the opinions of portfolio experts to develop a teaching portfolio prototype. *Studies in Education Evaluation, 30,* 305-321.

U.S. Department of Education (2001). *No child left behind act.* Retrieved from http://www2.ed.gov/policy/elsec/leg/esea02/107-110.pdf

U.S. News (2012). National University Rankings. *U.S. News and World Report*. Downloaded March 22, 2012 from http://colleges.usnews.rankingsandreviews.com/best-colleges.

Video Assessment of Interactions and Learning: VAIL (2010). Coding Manual (Unpublished Manual). Charlottesville, VA: Center for the Advanced Study of Teaching and Learning.

Veenman, M.V.J., Wilhelm, P. & Beishuizen, J.J. (2004). The relation between intellectual and metacognitive skills from a developmental perspective. *Learning and Instruction, 14*, 89-109.

Weisberg, D. Sexton, S., Mulhern, J., & Keeling, D. (2009). *The widget effect: Our national failure to acknowledge and act on difference in teacher effectiveness, 2nd Edition.* New York, NY: The New Teacher Project.

Wharton-McDonald, R., Pressley, M., & Hampston, J.M. (1998). Literacy instruction in nine first grade classrooms: Teacher characteristics and student achievement. *The Elementary School Journal, 99,* 101-128.

Wiens, P.D., (under review) Using a participant pool to gather data in a teacher education program: The course of one school's efforts.

Wiens, P. & Hessberg, K. (2011, December) *Do You See What I See? Using Video-Based Assessment to Measure Preservice Teachers' Ability to Recognize Effective Teaching.* Paper presented at the College and University Faculty of the National Council for the Social Sciences Annual Conference, Washington, DC.

Wiens, P.D., Hessberg, K., LoCasale-Crouch, J, & DeCoster, J. (2013). Using a standardized video-based assessment in a university teacher education program to examine preservice teachers knowledge related to effective teaching. *Teaching and Teacher Education, 33,* 24-33.

Wilson, S.M., Floden, R.E., & Ferrini-Mundy, J. (2002). Teacher preparation research: An insider's view from the outside. *Journal of Teacher Education*. 53: 190-204.

Wolf, K. & Dietz, M. (1998). Teaching portfolios: Purposes and possibilities. *Teacher Education Quarterly, 25,* 9-22.

Zeichner, K., (2005). A research agenda for teacher education. In M. Cochran-Smith & K. M. Zeichner, (Eds.). *Studying teacher education: The report of the AERA Panel on Research and Teacher Education.* (pp. 737-759). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Zeichner, K. (2012). Turn once again toward practice-based teacher education. *Journal of Teacher Education, 63,* 376-382.

Zeichner, K. & Conklin, H.G. (2008). Teacher education programs as sites for teacher preparation. In M. Cochran-Smith, S. Feiman-Nemser, D.J. McIntyre, & K.E. Demers, (Eds.) *Handbook of Research in Teacher Education* (3rd Edition, pp. 269-289). New York, NY: Routledge.

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

Zimmerman, R.S., Koury, E.L., Vega, W.A., Gil, A.G., & Warheit, G.J. (1995). Teacher and parent perceptions of behavior problems among a sample of African American, Hispanic, and non-Hispanic white students. *American Journal of Community Psychology, 23,* 181-197.

Zohar, A. (1999). Teachers' metacognitive knowledge and the instruction of higher order thinking. *Teaching and Teacher Education, 15*, 413-429.

Zohar, A., & Dori, Y.J. (2003). Higher order thinking skills and low-achieving students: Are they mutually exclusive? *The Journal of the Learning Sciences, 12*(2), 145-181.

Zucker, S., Christensen, R., Ellis, R.T., Harris, H., & Manning, D. (2004). *Using augmented norm-referenced assessments for NCLB compliance.* San Antonio, TX: Pearson. downloaded from http://www.pearsonassessments.com/pai/ai/research/resources/tmrsresources?squery=stanford

Zumwalt, K. & Craig, E. (2005). Teachers' characteristics: Research on the demographic profile. In M. Cochran-Smith & K.M. Zeichner, (Eds.). *Studying teacher education: The report of the AERA Panel on Research and Teacher Education.* (pp. 111-156). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.