Western Kentucky University
TopSCHOLAR®

Masters Theses & Specialist Projects | Graduate School

12-1-2005

The Residual Effect of Novice Primary Teachers on Reading Achievement Scores

Connie Mayo
Western Kentucky University

Follow this and additional works at: http://digitalcommons.wku.edu/theses

 Part of the Teacher Education and Professional Development Commons

Recommended Citation
Mayo, Connie, "The Residual Effect of Novice Primary Teachers on Reading Achievement Scores" (2005). *Masters Theses & Specialist Projects*. Paper 346.
http://digitalcommons.wku.edu/theses/346

This Dissertation is brought to you for free and open access by TopSCHOLAR®. It has been accepted for inclusion in Masters Theses & Specialist Projects by an authorized administrator of TopSCHOLAR®. For more information, please contact topscholar@wku.edu.

THE RESIDUAL EFFECT OF NOVICE PRIMARY TEACHERS ON READING ACHIEVEMENT SCORES

By

Connie Fort Mayo
B.S., Austin Peay State University, 1968
M.A., Austin Peay State University, 1974
Ed.S., Austin Peay State University, 2002

A Dissertation
Submitted to the Faculty of the
Graduate Schools of the University of Louisville
in Cooperation with Western Kentucky University
in Partial Fulfillment of the Requirements
for the Degree of

Doctor of Philosophy

Department of Leadership, Foundations, and Human Resource Education
University of Louisville and
Department of Educational Administration, Leadership & Research
Western Kentucky University

December 2005

THE RESIDUAL EFFECT OF NOVICE PRIMARY TEACHERS ON READING ACHIEVEMENT SCORES

By

Connie Fort Mayo
B.S., Austin Peay State University, 1968
M.A., Austin Peay State University, 1974
Ed.S., Austin Peay State University, 2002

A Dissertation Approved on

August 22, 2005

by the following Dissertation Committee:

Dissertation Director

DEDICATION

This dissertation is dedicated to my sisters and brother-in-law

Mary Beth Hudspeth

and

Susanne and Elvis Buttrey

whose selfless love and assistance have made this work and everything else in my life possible.

ACKNOWLEDGMENTS

I am grateful to my dissertation chair, Dr. Jeanne Fiene who taught me patience, humility, and persistence. She never gave up on me, or allowed me to give up on myself. She nurtured when nurturing was necessary and pushed me from the nest as I matured in the process. I also thank Dr. Joseph Petrosko, who was always available when I sought statistical insight. His advice was more than advice, it was the "how to" that completed Chapter III.

Dr. Thomas Reio, Jr., Dr. Jim Stone, and Dr. William Schlinker will forever have my gratitude for forcing me to look at my work with a more critical eye. Their recommended revisions, additions, and corrections during the writing process were invaluable.

I would be remiss if I did not express my appreciation for Dr. Ric Keaster who diligently read every word. His suggested corrections in style, punctuation, and word usage greatly improved this work. His expertise, time, and effort will not be forgotten.

I also want to thank my two wonderful sons who never once intimated that I was too old for this giant undertaking. Instead, they urged me to aim higher and accomplish more.

The most special "thank you" is directed toward my friend, study-partner, and fellow graduate, Dr. Brent VanMeter. Brent was only an e-mail

or a phone call away, day or night, with just the right word I needed in a sentence, the right word I needed for encouragement, or the right word I needed for my sanity during this painstaking process. My mainstay and my competition, he was always there for me in one of these roles, providing collegiality, support, and the impetus to continue the process.

ABSTRACT

THE RESIDUAL EFFECT OF NOVICE PRIMARY TEACHERS
ON READING ACHIEVEMENT SCORES

Connie Fort Mayo

August 22, 2005

This dissertation was an exploratory investigation of the residual effect of novice primary teachers on reading achievement scores. The study employed the theory of pedagogical content knowledge as a basis for understanding teacher expertise and comparing the expert teacher to the novice teacher. The research sought to answer two major questions; (a) Is there a statistically significant difference between the reading achievement, (measured two years later) of students taught by teachers of differing experience levels in primary grades and the state mean for the appropriate grade level? and (b) Is there a statistically significant difference among the two-year-later reading achievement of the groups of students based on teacher experience levels?

This study used student reading achievement test scores from the CTBS/5 Achievement Test published by CTB/McGraw Hill. Data were analyzed using the Statistical Package for the Social Sciences (SPSS) using a series of nine *t*-tests and three analysis of variance tests (ANOVAs).

While the findings of this study indicated that there were no statistically significant differences among the groups, the author discussed several limitations to the

study. In addition, proposals for future research in the area were presented. The final pages of the dissertation posit that school system administrators must use the information on novice teachers available to them to implement and strengthen programs of teacher recruitment, placement, training, and retention.

ABSTRACT

THE RESIDUAL EFFECT OF NOVICE PRIMARY TEACHERS ON READING ACHIEVEMENT SCORES

Connie Fort Mayo

August 22, 2005

This dissertation was an exploratory investigation of the residual effect of novice primary teachers on reading achievement scores. The study employed the theory of pedagogical content knowledge as a basis for understanding teacher expertise and comparing the expert teacher to the novice teacher. The research sought to answer two major questions; (a) Is there a statistically significant difference between the reading achievement, (measured two years later) of students taught by teachers of differing experience levels in primary grades and the state mean for the appropriate grade level? and (b) Is there a statistically significant difference among the two-year-later reading achievement of the groups of students based on teacher experience levels?

This study used student reading achievement test scores from the CTBS/5 Achievement Test published by CTB/McGraw Hill. Data were

analyzed using the Statistical Package for the Social Sciences (SPSS) using a series of nine *t*-tests and three analysis of variance tests (ANOVAs).

While the findings of this study indicated that there were no statistically significant differences among the groups, the author discussed several limitations to the study. In addition, proposals for future research in the area were presented. The final pages of the dissertation posit that school system administrators must use the information on novice teachers available to them to implement and strengthen programs of teacher recruitment, placement, training, and retention.

TABLE OF CONTENTS

PAGE

DEDICATION iii
ACKNOWLEDGMENTS. iv
ABSTRACT v
LIST OF TABLES. xi

CHAPTER

I. INTRODUCTION. 1

Background 3

Statement of the Problem 6

Relationship of this Study to the Problem 10

Purpose of the Study 11

Conceptual Model. 12

Research Questions 13

Null Hypotheses 14

Significance of the Study. 14

Definition of Terms 15

Organization of the Study. 16

II. REVIEW OF THE LITERATURE. 17

The Expert Teacher 22

The Novice Teacher 38
Comparing Novice to Expert 59
Relating Teacher Expertise to Student Achievement 89
Teacher Characteristics (Presage Variables) 89
Teacher Behaviors (Process Variables) 113
Presage and Process Variables 125
Summary and Research Problem 129
Research Purpose 132
III. METHODOLOGY 134
Subjects 142
Design 149
Instrumentation 158
Procedures 159
Data Analysis 163
Assumptions and Limitations 167
IV. RESULTS OF THE STUDY 168
Description of the Participants 170
Statistical Analysis 175
Summary of Findings 186
V. SUMMARY AND CONCLUSIONS 191
Discussion of Research Findings 192
Review of Literature 193

Research Design 193

Statistical Analysis. 194

Limitations and Cautions. 195

Implications. 196

Recommendations for Practitioners. 198

Suggestions for Future Research. 199

Conclusions 200

REFERENCES 202

APPENDIXES. 214

CURRICULUM VITAE. 218

LIST OF TABLES

TABLE PAGE

1. Identified Characteristics of Expert Teachers. 36

2. Identified Characteristics of Novice Teachers. 57

3. Comparing Novice Teachers to Expert Teachers 86

4. Relating Teacher Characteristics to Student Achievement 112

5. Relating Teacher Behaviors to Student Achievement. 123

6. Relating Presage and Process Variables to Student Achievement. . .129

7. Fourth Grade Mean Reading Scale Score Gain (2002) 137

8. Tennessee Grading Scale for Fourth Grade Student Gains (2002). . . 138

9. Normal Curve Equivalet (NCE) Scores for Experience Level Groups in Grades Two, Three, and Four. 139

10. Tennessee Grading Scale for Student Achievement 140

11. Classroom Variables 145

12. Description and Estimated Numbers of Students in Two-year-Later Comparison 149

13. One Sample *t*-test Comparisons to Address Research Question 1 . . .151

14. One Sample *t*-test Comparisons to Address Research Question 2 . . .152

15. One Sample *t*-test Comparisons to Address Research Question 3 . . .153

16. Kindergarten Data Collection Activities. 154

17. First Grade Data Collection Activities 155

18. Second Grade Data Collection Activities. 156

19. Independent Variable: Experience Level. 165

20. Descriptive Analysis of Overall Student Sample 174

21. Group Mean for Sample Groups Two Years Later. 176

22. Student Scores of Novice Primary Teachers Two Years Later (*t*-tests 1-3) Compared to State Mean (52 NCE). 177

23. Student Scores of Second and Third Year Primary Teachers Two Years Later (*t*-tests 4-6) Compared to State Mean (52 NCE . . . 178

24. Student Scores of Fourth or More Year Primary Teachers Two Years Later (*t*-tests 7-9) Compared to State Mean (52 NCE). 180

25. Scores of Students with Kindergarten Teachers at Three Levels of Experience Two Years Later Compared to State Mean (52 NCE) . . . 181

26. Scores of Students with First-Grade Teachers at Three Levels of Experience Two Years Later Compared to State Mean (52 NCE). . . 182

27. Scores of Students with Second-Grade Teachers at Three Levels of Experience Two Years Later Compared to State Mean (52 NCE) . . . 183

28. Analysis of Variance for Reading NCEs by Experience Level of Kindergarten Teacher (ANOVA – 1). 184

29. Analysis of Variance for Reading NCEs by Experience Level of First Grade Teacher (ANOVA – 2). 185

30. Analysis of Variance for Reading NCEs by Experience Level of Second Grade Teacher (ANOVA – 3). 186

31. Decisions on Null Hypothesis for each *t*-test. 188

32. Decisions on Null Hypotheses for each ANOVA 189

CHAPTER I

"Teaching is knowing . . ." (Harrington, 1994, p. 191). This knowing encompasses several knowledge components including: (a) pedagogical knowledge, (b) personal beliefs and practical experience, (c) knowledge of learners and learning, (d) subject matter knowledge, (e) contextual knowledge, and (f) pedagogical content knowledge (Carlsen, 1999; Magnusson, Krajcik, & Borko, 1999; Morine-Dershimer & Kent, 1999). When coupled with experience, these types of knowledge develop and mature into a complex quality that surpasses mere subject matter and teaching techniques. Beijaard and Verloop (1996) term this complexity the "core of teaching quality."

Recent legislative developments have focused unprecedented attention on the quality and knowledge of the nation's teaching force. The mandate of No Child Left Behind (PL 107-110) (NCLB) (United States Department of Education, n.d.) for a "highly qualified teacher" in every classroom was precipitated by a growing body of research identifying the classroom teacher as a critical factor in student learning. Ferguson (1991) found that teacher expertise accounted for about 40% of the variation in students' reading and math achievement. Darling-Hammond (2000) found that student

achievement appeared to be more strongly related to teacher quality than class sizes, overall spending levels or teacher salaries. The National Commission on Teaching and America's Future (1996) determined that ". . . the school reform movement has ignored the obvious: What teachers know and can do makes the crucial difference in what children learn" (p. 5).

Leslie Huling, chair of the Texas State Board for Educator Certification's Panel on Novice Teacher Induction Support System asked readers to think how they might react to being told their children had been selected to be placed only in the classrooms of novice teachers for their entire K-12 career. She suggested that most educators and parents know intuitively that teacher quality improves with experience (Texas State Board, 1998). Huling continued by questioning whether a child's success or failure in school could be attributable to the circumstance of placement.

In Tennessee, the novice teacher has not yet completed the certification process and is still teaching on an apprentice license. Full licensure is granted at the end of the third year of teaching. This three-year apprentice period grants the beginning teacher time and opportunities to initiate, implement, and institutionalize effective teaching characteristics and behaviors.

The first year in the classroom is a critical stage in the apprenticeship. It is characterized by the transition from 17-year student to novice teacher with tensions emanating from exploration, challenge, and expansion. Success

in this phase requires courage for risk-taking and the ability to learn from experience. Learning to teach is not always a comfortable process. The willingness to take risks by experimenting with new strategies and behaviors that break with traditional structures allows beginning teachers to test and be comfortable with their teaching styles and strengths. In addition to learning about themselves and their abilities, first-year teachers must teach the state standards and prepare students for success on the spring achievement test and their next year in school.

Background

In a search for educational quality, England created a payment-by-results program in the eighteenth century. The purpose was to improve English education by paying teachers according to their students' performance on examinations. England's 1710 experiment in basing teacher pay on student scores in reading, writing, and arithmetic failed due to the swift response of the schools to limit their curriculum to the measured basics and to manufacture test results. Graduating students had mastered little of what was supposedly taught. This early attempt at identifying and rewarding quality education was just one of dozens that populate the education history landscape (Wilms & Chapleau, 1999).

Today, most educational policymaking has its origin in special interest groups and is adopted by either state or federal government. Secretaries of Education Bell, Bennett, and Riley maintained The "Wall Chart," comparing

state level educational data, for many years. This chart was most likely the predecessor of Congress's 1990 reorganization and reauthorization of the National Assessment of Educational Progress (NAEP) allowing for comparisons of student achievement levels by state (Vinovskis, 1999). In turn, NAEP data provided statistical substance to the current concerns about educational accountability.

State legislatures rushed to establish accountability plans, which addressed this national focus. By 2001, all 50 states had mandated testing programs. Each state established accountability departments or offices of teacher quality within their respective education system and implemented a method of reporting progress to the public (Goetz & Duffy, 2001).

Tennessee was no exception. In 1992, the Tennessee General Assembly enacted the Education Improvement Act. This act set five performance standards for each school in the state. These addressed two academic standards and three non-academic standards. Academic standards were based on achievement and value-added test data. Non-academic standards focused on attendance rate, dropout rate, and promotion rate (State of Tennessee, 1996).

To obtain the necessary academic data, Tennessee's Education Improvement Act provided for the annual administration of an achievement test to all students in grades three through eight. The act required that the specific test used for this assessment be bid for purchase every five years. The

test currently in use is the TerraNova, published by CTB/McGraw Hill. Results of this test supply the state with information for reporting school and system progress to the schools and the public.

To analyze the data gathered from this annual assessment, members of the legislature selected Dr. William Sanders' model for test data interpretation. This model, the Tennessee Value-Added Assessment System (TVAAS) has become the guiding force for instructional decisions made in Tennessee school systems and individual schools since that date. Sanders, Saxton, and Horn (1997) contended student achievement was not a valid measure by itself, because it allowed students to remain at a fixed percentile score and make no growth toward improving that score. The premise of the Sanders Model is that a student, or group of students, should make one year's progress, regardless of their beginning scores.

For example, end-of-year fourth grade reading students were expected to have a scale score gain of 12 points over the preceding spring when they were tested as third graders. This is true whether they began with a scale score of 300 and progressed to 312, or they began with a scale score of 500 and progressed to 512. Either set of scores would equal one year of reading progress. In 2004, Sanders revised his model to use NCE scores on a criterion-referenced test instead of scale scores on a norm-referenced test. Currently, students and teachers are evaluated for growth based on a normal curve equivalency (NCE) score gain of zero or above to show a year of value added.

Reports promulgated by this formula are the basis for data-driven goals required by both the Tennessee School Improvement Plans and the Southern Association of Colleges and Schools (SACS). In addition, the State Department of Education publishes annual school and school system report cards, which give each grade group at each school a grade in achievement (grades three through five and six through eight). These grades are stated as an A, B, C, D, or F in each subject area. As a major part of the annual report, all teachers in those grades receive a personal report disclosing whether the students they taught the previous school year made a year's gain in each subject. Each teacher is expected to make a minimum of a one-year gain.

Statement of the Problem

One component of the Tennessee accountability program is a goal for the students of each system, school, grade level, and teacher to make a year's growth. This component is titled Value Added Assessment. Progress toward this goal is determined by measuring NCE gains on an annual achievement test. These gains are calculated by comparing the NCEs of students for two consecutive years. An NCE gain of zero earns the letter grade of "C." For example: If the students' mean NCE score in the fourth grade in School 1 in 2004 is 56, the score is compared to the scores of these same students last year (in the third grade) which was also 56. School 1 receives a letter grade of "C" in value added. Other letter grades are assigned depending on the improvement of NCE scores.

A Tennessee school system, with system-wide and school achievement test percentile scores at-or-above the national norm (50 percentile), has never met the expected 100% of the expected NCE gain goal in all subjects and all grade levels to date. The system has adopted a standards-based and test-objective-driven curriculum, enhanced professional development, and increased expenditures for technology and other innovative programs. These interventions have not provided the impetus needed to meet the gain goal. During the 15-year data-gathering period from 1990 through 2004, no school in this system, neither grade level, nor subject area has been consistently excellent, fair, or poor.

Due to the complexity of the education process, several variables have the potential to cause the variation in scores. These include poverty level, ethnic diversity, textbook differences, local funding for instructional materials, classroom schedules, and teacher degree or certification. System-wide, all of these variables were equal, equitable, or comparable.

Another widely accepted variable is the classroom teacher. Many specific teacher characteristics create variance. Organizational skills, flexibility, classroom management, instructional strategies, classroom environments, and the ability to differentiate instruction vary greatly. The reviewed literature illustrates that novice teachers are struggling to gain competence in these areas. Despite the differences between the two groups, novice teachers are assigned the same numbers of instructional objectives to

teach the same numbers of students, in the same period as the more experienced teachers. The consistency of these expectations is in direct contrast to the variances in their level of preparation.

Thirty-three percent of this system's educators have taught fewer than three years in the system. The majority of these teachers are within the first three years of their profession. Does this large percentage of novice teachers have an effect on the NCE gain inconsistencies in this school system?

Accountability is a natural concern that has promulgated each individual state's respective action. This concern eventually focuses on the classroom teacher's locus of control. Classroom teachers must structure instruction, foster achievement, and ensure gains for students in their charge. The individual classroom teacher is still the single most important piece in the entire educational jigsaw puzzle.

However, if all teachers, schools, and systems are to be accountable to the same standards, more attention must be paid to the differences, which make each of them unique. Of particular importance is the list of characteristics, which separate expert teachers from others.

Expertise is an elusive concept in the complex task of teaching. It is often defined in terms of student achievement and just as often by the subjectivity of reputation. The growing body of research on effective teaching reinforces the idea that expert teachers can be identified by specific characteristics, behaviors, and qualities. These elements include teacher

preparation, classroom management, planning, instruction, assessment, and the personal qualities that govern interactions with students (Stronge, 2002). These qualities often correlate and are associated with teacher experience (Leinhardt & Greeno, 1986; O'Connor & Fish, 1998; Schempp, Manross, Tan, & Fincher, 1998; Stronge, 2002; van Driel, Verloop, & de Vos, 1998)

Regardless of the quality of their preservice program, learning to teach inevitably occurs after teachers begin to teach. The first year of teaching is an intense and potentially formative phase in learning to teach (Nemser, 1983). The process is not mastered in one year. Berliner (1986) identified stages in learning to teach. The first stage, survival, is characterized by concern about coping, a focus on themselves and their needs, lack of perceived responsibility for student achievement, classroom management problems, reactions rather than proactions, lack of flexibility, and little differentiation (Katz, 1972). Moir (1990) characterized the first-year survival period as one of disillusionment, rejuvenation, reflection, and anticipation. Novice teachers are beginning to craft a professional identity through their struggles and explorations of students and subject matter (Feiman-Nemser, 2001)

Novice teachers' emotions during this critical time include fear, exhilaration, loneliness, confusion, and frustration (Zepeda & Mayers, 2001). Veenman (1984) termed this phenomenon as reality shock and referred to it as the collapse of the missionary ideals formed during teacher training. Ganser (1997) said the first year of teaching ". . . is like being in water over

your head. You are floating on a tiny piece of foam that crumbles away every day just a little bit" (p. 106).

The problems faced by first-year teachers include classroom management (Stanulis, Campbell, and Hicks, 2002; Ward & O'Sullivan, 1998), isolation (Cleary & Groer, 1994; Westerman, 1991), and frustration (Bullough, 1987; Veenman, 1984). They enter the classroom armed with "book" knowledge of subject matter, a few strategies they learned during student teaching, and no experience in planning for student mastery. These novices are not equipped for the responsibilities of an isolated workplace where they seldom hear or see each other teach (Rosenholtz, 1989).

The number of novice teachers sharing these characteristics will increase greatly during the next decade due to growing student populations and an aging teaching force. Student enrollment is predicted to climb from 50 million in 1995 to 54.3 million by 2007. Additionally, one fourth of the current teaching force were found to be 50 years old or older in 1997 (Darling-Hammond, 1997). The literature review examined the role of experience in teacher characteristics, behaviors, and student achievement. The study determined if there were residual or long-term differences among students taught by teachers with differing experience levels.

Relationship of the Study to this Problem

The push for accountability forced the faculty of each Tennessee school into self-examination for strengths and weaknesses. In an otherwise

successful school system, inability to meet minimum goals in all academic areas and grade groups raises the question of probable cause. A characteristic that is both obvious and disturbing to an administrator establishing a stable team of educators is the high turnover rate, which contributes to a steady population of novice teachers and those with only minimal experience.

There is a growing body of research on the characteristics of both novice and expert teachers, the differences between the two groups, and how expertise affects student achievement. There is, however, little research that addresses the idea that future success in school is directly related to the expertise of primary teachers; and more specifically, the residual effects of reading instruction in the primary grades.

Purpose of the Study

The purpose of this descriptive study was to analyze the residual or persistent differences among reading achievement of students taught by differing levels of teaching experience. The researcher investigated the effect of teachers' years of teaching experience on student achievement two years later. Reading was chosen as the subject area for the study because most academic areas display some dependence on success in this skill. As schools across the nation implement NCLB and pursue the goal of closing the achievement gap, literacy and reading instruction are keys to achieving those objectives. Characteristics of teachers who deliver the instruction in these areas are vitally important.

Conceptual Model

The conceptual model to support this study is illustrated in Figure 1. The model demonstrates that the study determined if there is a significant difference in the TerraNova Reading Achievement Test scores (two years later) of students served by novice, relatively inexperienced, and experienced teachers. Scores of all three groups were compared to the state mean NCE scores and to each other. Additionally, the study determined if there is a statistically significant difference among the scores of teachers' students, from one year to the next, during their first three years of teaching.

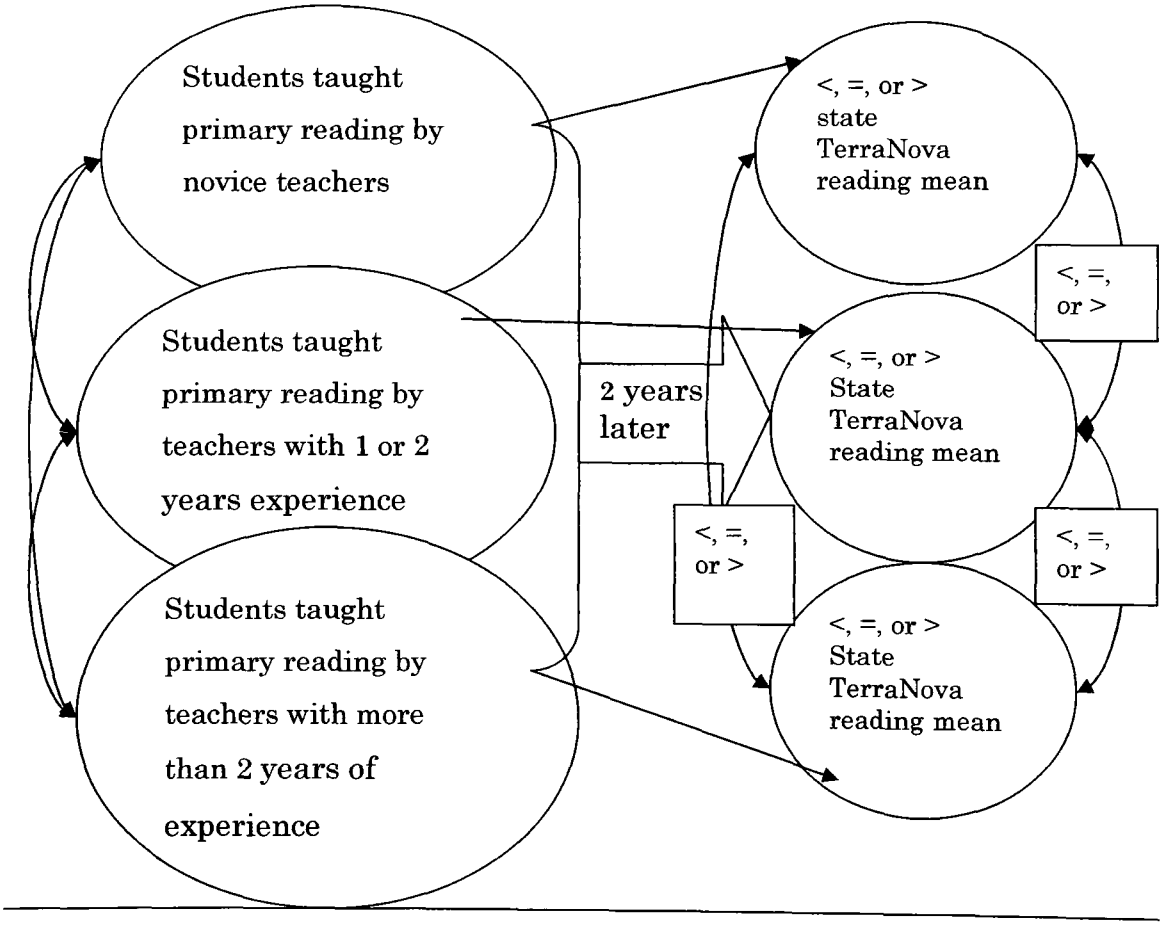


Figure 1. Conceptual model to support study questions two through four

Research Questions

Tennessee assessment data from 2000 through 2004 were collected and analyzed to determine if there was a significant difference in students' two-year-later reading achievement scores among students whose primary grades teachers had differing levels of experience. The data were analyzed to answer the following questions:

1. Is there a statistically significant difference between the reading achievement, measured two years later, of students taught by a novice teacher in primary grades (kindergarten, one, and two) and the state mean for the appropriate grade level?
2. Is there a statistically significant difference between the reading achievement, measured two years later, of students taught reading by a second or third year teacher in primary grades (kindergarten, one, and two) and the state mean for the appropriate grade level?
3. Is there a statistically significant difference between the reading achievement, measured two years later, of students taught reading by a teacher with three or more years of teaching experience (teachers in their fourth or greater year of teaching) in primary grades (kindergarten, one, and two) and the state mean for the appropriate grade level?

4. Is there a statistically significant difference among the two-year-later reading achievement of the three groups of students in questions one, two, and three?

Null Hypotheses

Null Hypothesis 1:

There will be no significant differences between the two-year-later reading NCE scores among students whose primary reading teachers have differing levels of experience and the mean reading NCE for the state.

Null Hypothesis 2:

There will be no significant differences in the two-year-later reading NCE scores among students whose primary reading teachers have differing levels of experience.

Significance of the Study

Knowledge gained from this study can be beneficial to board members and system-wide personnel as they plan future recruiting, placement, and teacher retention strategies through mentor programs and other scaffolding strategies for teacher induction. Additionally, school administrators can benefit as they make placement decisions for students. Most importantly, students should benefit by the highest quality instruction available during the vital first years of learning to read.

This study also contributed to the growing body of knowledge on teacher development, effective teachers, and the effect they impose on

student achievement. In this era of increased accountability for classroom teachers, this study provided insight into the variables that characterize and affect their effectiveness and, in turn, student achievement

Definition of Terms

The following list of definitions ensures clarity and understanding throughout the study. The researcher composed all definitions with the exception of those with a citation.

1. Accountability: a systematic collection, analysis, and use of information to hold schools, educators, and others responsible for the performance of students and the education system (Education Commission of the States, 1998)
2. Adequate yearly progress (AYP): a measure of a school's or school system's ability to meet federal benchmarks with specific performance standards from year to year (Tennessee Department, n.d.)
3. NCE Gain: The difference between a student's or group of students' NCE scores on a norm referenced test from one year to the next
4. Regular classroom teacher: a teacher assigned to teach a heterogeneous group of students
5. Value-added: measurement of student progress within a grade and subject, which demonstrates the influence the school and teacher have on the students' performance. This reporting provides diagnostic

information for improving educational opportunities for students at all achievement levels.

Organization of the Study

Chapter I includes the study introduction, background, statement of the problem, purpose of the study, conceptual model, research questions, hypotheses, significance of the study, definition of terms, and organization of the study. Chapter II contains the literature review and related research. In Chapter III the methodology and data-gathering procedures for the study are delineated. The results of the data analysis are presented in Chapter IV. In Chapter V, the researcher summarizes the study and findings, draws conclusions from the findings, and discusses recommendations for further research.

CHAPTER II

REVIEW OF THE LITERATURE

Is there a connection between chess, sports, and teaching? When and how is the evolution made from novice to expert in these three endeavors? Since the late 1800's when Alfred Binet (Cunningham, 1995) piqued interest in the study of expertise by studying mental calculators and chess players, others have followed with studies of expertise in various fields. These include de Groot's (1965) study of chess players' thinking. de Groot found that chess masters recognized meaningful chess configurations and realized the strategic implications of situations. This allowed them to consider sets of possible moves or actions that were superior to others. de Groot concluded that increasing experience and knowledge in a specific field has the effect that things, which at earlier stages had to be abstracted or even inferred, are apt to be immediately perceived at later stages.

Chi, Feltovich and Glaser's (1981) research on expertise in physics resulted in findings that agreed with de Groot. They found that experts have multiple ways of thinking about the domain and many different representations of the knowledge domain. They can switch from one representation to another, and they have the meta-knowledge that allows

them to know which representation to choose for which task and which representation to switch to while solving the task.

Rosenbloom and Newell's (1986) work resulted in the power law of learning. This law states that with more practice at a task, people seem to always be getting faster. However, the rate of learning decreases the more practice one has. Studies in the fields of medicine, sports, music, and mathematics later contributed to the field.

Gobet's (n.d.) summarized findings from these and other studies include (a) it takes years to become an expert (Chase & Simon, 1973), (b) experts showed the same cognitive limits (Simon, 1979), (c) experts did not generally have a higher intelligence than non-experts, and no special talent, but acquire expertise through deliberate practice and training (Ericsson & Charness, 1994), (d) experts performed intuitively and could not verbalize their expertise (Gobet, n.d.).

Practice is an integral part of many of the findings on expertise. Experience in teaching, or any field, provides practice. Lesgold, Rubinson, Geltovich, Glaser, Klopfer, and Wang (1988) determined that even though experience is often associated with and typically developed as a function of expertise, experience is not synonymous with expertise. Ericsson and Charness (1994) contended that individuals who seek to become expert undergo long periods of active learning, refining and improving their skills and performance under the tutelage of an expert teacher or coach. They also

found that experts make a significant investment in learning all they can in their area and enjoy talking about their field, gathering others' views on pertinent topics.

Bransford, Brown, and Cocking (1999) identified six key principles of experts' knowledge and their potential implications for learning and instruction:

1. Experts notice features and meaningful patterns of information that are not noticed by novices.
2. Experts have acquired a great deal of content knowledge that is organized in ways that reflect a deep understanding of their subject matter.
3. Experts' knowledge cannot be reduced to sets of isolated facts or propositions but, instead, reflects contexts of applicability: that is, the knowledge is "conditionalized" on a set of circumstances.
4. Experts are able to flexibly retrieve important aspects of their knowledge with little attentional effort.
5. Though experts know their disciplines thoroughly, this does not guarantee that they are able to teach others.
6. Experts have varying levels of flexibility in their approach to new situations. (pp17-18)

The research in expertise in various fields offers the educational community considerable insight into teaching expertise. Not everyone will reach the pinnacle of expert teacher. However, when educators are aware of the elements of expertise in the field of teaching they can become more expert. In this chapter, the researcher will review literature relevant to the acquisition, importance, and effects of expertise in the field of teaching.

The first section contains a literature review on the attributes of expert teachers. Especially relevant to this section is the concept of pedagogical content knowledge (PCK). This concept of PCK was conceived by

Shulman (1986) and theorizes that effective teachers have a special kind of pedagogical knowledge and content knowledge that they meld into their teaching.

> (PCK includes) the most useful forms of representation of [topics], the most powerful analogies, illustrations, examples, explanations, and demonstrations –in a word, the ways of representing and formulating the subject that make it comprehensible to others. . . Pedagogical content knowledge also includes an understanding of what makes the learning of specific topics easy or difficult: the conceptions and preconceptions that students of different ages and backgrounds bring with them to the learning of those most frequently taught topics and lessons. (p. 9)

Although the concept of PCK is not universally accepted, it provides a basis for much of the research on exploring the knowledge that effective teachers possess and how they teach.

Studies included in section two center on novice teachers and the forces that shape, not only their first year, but the years to come. These forces include beliefs, skills, knowledge, strategies, experiences, and growth. Novice teachers enter their teacher preparation courses with a set of dispositions about education formed by 13 years of experiencing education as a student. Some of these beliefs persist through their teacher education courses and some change. When they enter the educational field as teachers, these beliefs are again challenged by their new environment. As novice teachers enter this new environment they must develop *skills* in classroom management, communication with new peers, and adjusting to the culture of the institution.

Additionally, they are confronted by the requirement to be knowledgeable in their field of study, whereas in the past they have filled the role of learner. This new role is a source of stress. Novice teachers must begin to develop a repertoire of *strategies* to differentiate their teaching for the variety of students they will be serving. First-year teachers will encounter experiences they were not prepared for at the university level. They will be responsible for discipline, parent conferences, evaluation of teaching materials, and a set of learning standards (often from another state). Lastly, they will grow and mature as educators, both inside the classroom as they teach and outside the classroom as they prepare for instruction. Section two will provide insight into how these forces contribute to the development of first year teachers.

The third section consists of reviews which highlight research on the specific differences between novice and expert teachers and what implications these differences have for classroom instruction. Again, these studies are generally qualitative in nature. Most are small in terms of sample size and define both characteristics and behaviors which characterize the expert teachers. As in the first two sections, all the studies identify experience as either a variable or a finding.

The final section includes reviews of research studies which define the connection between expert teachers and student achievement. Many of these studies are based on the production function model measuring input and its

relationship to student outcomes. They all lend credence to the idea that the teacher is the most important variable in any classroom. These studies illustrate that one characteristic of an expert teacher is years of reflective practice and this expertise is exemplified in the achievement of their students.

The Expert Teacher

The expert teacher exerts a mysterious pull on the minds of educational researchers. Her siren call entices them to spend innumerable hours observing her actions, questioning her mental deliberations, and probing through countless documents which might yield a clue to her ability to effect learning. According to Brophy and Good (1986), an effective teacher (a) plans lessons for student mastery, (b) has good classroom management, and (c) exhibits high expectations for her students. Does it stop there or are there unexplored depths to teacher expertise? What does the acquisition of this knowledge mean for children?

Modern research on teacher effects was conceived in the production function analysis research of Hanushek (1986). For at least 21 years, Hanushek published a vast array of studies comparing resources to student achievement. These resources included teacher experience and education because, historically, teacher salaries have been tied to these two characteristics. On the whole, he found no consistent or positive relationship to school spending and student achievement (Hanushek, 1996). While

Hanushek did find some positive correlation between teacher experience and student performance, he attributed that correlation to the more experienced teachers selecting teaching assignments in higher socio-economic schools (Hanushek, 1993). His findings have been widely accepted in academic, legal, and public policy arenas.

Interest in teacher expertise was further nurtured by the expertise research of Glaser and Chi (1988) who determined that expert knowledge is characterized as involving the development of organized conceptual structures, or schema, that guide how problems are represented and understood. These studies piqued the interest of others who sought to isolate the variables which contributed to teacher expertise. In the 1980s, the search for answers to questions about teacher expertise centered on the concepts of a few researchers. Berliner (1986), Brophy and Good (1986), and Shulman (1986) conceptualized teacher expertise into stages of development, attributes, and types of knowledge. Since that time, many additional researchers have used the concepts of these men to frame their investigations (Hastie & Vlaisavljevic, 1999; Leinhardt & Greeno, 1986; Schempp, Manross, Tan, & Fincher 1998).

In this section, literature that sheds light on these inquiries was reviewed. The foci of the studies were:

1. What goes on in expert teachers' classrooms?
2. What types of knowledge do they need to become an expert teacher?

3. How are these types of knowledge developed? and
4. Do all expert teachers have the same types of knowledge?

Leinhardt and Greeno's (1986) study elucidated the activity structures and routines of skilled teachers by describing them, analyzing their frequency and duration, analyzing the functions of routines, and contrasting activities of novice and expert teachers. Their grounded theory design resulted in the identification of models for ten activity structures utilized by expert teachers.

The researchers identified expert teacher participants by reviewing the growth scores of students over a five-year period and selecting the 15 teachers with the highest growth scores. Eight of these teachers agreed to participate. The authors did not specify the size or location of the larger group from which these teachers were selected. All taught in self-contained elementary classrooms.

Observations occurred over a three-and-a-half-month period and included approximately one-fourth of the mathematics classes taught during that period. Open-ended notes of the observations, pre- and post-observation interviews, and videotapes of three to five classes comprised the data sources. The notes were segmented into action records which listed durations, actions of students, and a name for the teacher's action. The researchers defined the actions and used the definitions for analyzing additional transcriptions or videotapes. Next they used the codable data to determine medians and ranges of time spent in each activity. Finally, they analyzed the videotapes

drawing on information from the stimulated recalls, interviews, and other transcribed discussions with the teachers.

Leinhardt and Greeno (1986) identified ten categories to describe the actions of expert teachers: (a) presentation and review, (b) shared presentation, (c) drill, (d) game drill, (e) homework, (f) guided practice, (g) monitored practice, (h) tutoring, (i) test, and (j) transition. They characterized the experts' lessons as action agendas consisting of a list of action segments. These agendas were with well-practiced routines with set patterns of the identified activities. This use of routines reduced the cognitive processing time for teachers allowing lessons to flow without interruption. They characterized teaching as a complex cognitive skill determined, in part, by the nature of teacher's knowledge system.

This expert teacher finding reflected the earlier work of Glaser and Chi (1988) who determined that experts possess superior self-monitoring and self-regulating skills allowing performance to appear automated. The researchers concluded by explaining that this analysis of routines and activity segments was an early step in understanding how expert teachers organize for successful instruction.

The ten teacher routines and actions identified by Leinhardt and Greeno (1986) begin to comprise the pedagogical aspect of pedagogical content knowledge (PCK). Expert teachers meld these activities with appropriate content.

Schempp, Manross, Tan, and Fincher (1998) studied the influence of subject matter expertise on PCK of physical education teachers. Their qualitative study assessed teachers' subject area expertise in three ways: (a) self-reported subject knowledge expertise rating scale, (b) participation in the area of expertise for a sustained period of time and in a variety of capacities (e.g., coach, referee), and (c) background interview to probe depth and breadth of subject matter knowledge. Schempp et al. grounded this research in the Berliner's (1988) theory of teacher development.

Using the recommendation or reputation method, Schempp et al. (1998) selected 10 elementary or middle public school teachers and student teachers as participants in their study (five novice and five competent). (The authors did not cite the source for their selection method.) Each of the five competent teachers (a) had five or more years of teaching experience, (b) received a recommendation by his or her peers or university faculty, and (c) sustained acceptable service as a cooperating or mentor teacher. The 10 study participants averaged eight years of experience teaching physical education in public middle schools (N= 10).

The self-reported rating scale allowed these teachers to rate their expertise in 25 areas identifying areas in which they were most expert. Four interviews completed the data collection process. Each interview gathered specific information from the participants. The first interview garnered background information on the participant including subject matter

knowledge and experience. The second involved the participant planning two hypothetical units (one in the area of expertise and the other in an area of non-expertise). In the third interview, the participants each planned a lesson in their area of expertise and a lesson in an area of non-expertise. The fourth interview involved retrospection. Two investigators each conducted a one-hour interview as a team.

Schempp et al. (1998) audiotaped and transcribed the 40 interviews. Using constant comparative analysis and working independently from each other, they developed themes and categories. They then divided the themes and categories into expert and non-expert subject areas. Finally, they compared individual findings for uniformities or trends identifying specific manifestations of teachers' knowledge as influenced by their subject expertise. Schempp et al. reported their findings by similarities and differences.

In general, the study findings suggest a significant difference in the teaching skills of teachers teaching in their subject of expertise compared to the same teachers teaching in areas of non-expertise. These differences included (a) recognizing problems in student learning, (b) level of detail in planning and organizing subject matter, (c) accommodating a wider range of student abilities and skills, and (d) comfort with and enthusiasm for teaching the skill. The study findings reflected those of Hashweh who contended subject experts were more able to transfer their expertise into pedagogical

activities than non-experts (as cited in Schempp et al., 1998). The researchers suggested that teacher education and in-service programs stress the acquisition of subject matter expertise. They contended this might enable teachers to become more effective and enthusiastic teachers.

In a similar study, Hastie and Vlaisavljevic (1999) examined the relationship between teachers' self-reported subject matter expertise and their presentation of instructional tasks and accompanying levels of accountability. Their study was framed by the ecological model of classroom dynamics with a focus on the subject matter arm of the model. The authors felt this case study, in a variety of physical education content areas, might provide some insight into how subject matter might be related to the work students do and how they organize that work.

Nine high school physical education teachers comprised the participant sample. Their experience level ranged from 6 to nearly 40 years. First, Hastie and Vlaisavljevic (1999) assessed each participant's level of subject matter expertise (SME) by conducting structured interviews on the teacher's educational background and personal experiences. This method was used so that the teachers would commit to either *higher* or *lower* knowledge of a sport they would teach. The researchers then observed each teacher conduct two classes during the third week of a five-week unit for a total of 18 observations, (one in each level of SME). They created an observation outline to categorize the type of task teachers were performing and their level of

accountability. Tasks were categorized by (a) informing, (b) extending, (c) refining, and (d) applying. Accountability measures were categorized as either *less demanding* or *quality.* Next, the authors conducted a paired-sample *t*-test to compare task presentation and accountability for each teacher in each identified level of SME.

Teachers instructing in their higher SME area used significantly more tasks per lesson than lower SME teachers. Lower SME instruction relied heavily on informing and refining tasks often failing to use extension tasks. Higher SME teachers utilized application and extension more often. Lower SME teachers used degree of participation or effort for accountability measures, whereas the higher SME teachers looked at quality of student performance.

Although this study focused on subject matter expertise, there seemed to be a link to these findings and PCK findings by Grossman (1990). He suggested that teachers who have strong PCK formulated subject matter and presented it in understandable ways. Likewise, teachers with weak PCK struggled with lesson design, appropriate progesssions and successful monitoring of student performance (Rovegno, 1992).

O'Connor and Fish (1998) examined teacher experience from a different perspective, whether or not teacher experience influenced the classroom system by using a systems perspective. Systems theory

conceptualizes the interaction of all members of any system and how the various relationships influence that system.

This study employed classroom observations utilizing the *Classroom Systems Observation Scale* (CSOS). The CSOS was created to evaluate the interactions in a classroom on three dimensions: (a) level of flexibility, (b) cohesion, and (c) communication. The sample consisted of 35 novice and 35 experienced elementary teachers representing 18 schools in New York State (ten private and eight public schools). Novice teachers had less than one year of experience. Principals recommended the experienced teachers because of their exceptional teaching ability. Each experienced teacher had five years or more teaching experience and one or more at their present grade level. The average experience level was 18 years and the range was 5 to 43 years.

It appeared that 20 of the participants (10 novice and 10 experienced) participated in the pilot study only. The researchers conducted 50-minute observations on the remainder of the participants using the CSOS instrument. Next, they employed a series of *t*-tests to determine if there was a difference in the levels of flexibility, cohesion, or communication (as measured by the CSOS) between the two groups of teachers ($N = 20$).

O'Connor and Fish (1998) found that the classrooms of experienced teachers were significantly more flexible than those of the novice teacher. They ascertained that experienced teachers' classrooms showed a

significantly higher level of communication. The cohesion dimension was not significantly different between the two cohorts.

From a systems perspective, this study showed teachers' experience to be positively related to their ability to be responsive and adaptable (flexible). Borko and Livingston (1989) and Westerman (1991) referred to this concept as interactive teaching. The communication dimension also showed a positive correlation to years of teaching experience with teachers using students' questions and responses to guide discussion as described by Cleary and Groer (1994). In the dimension of cohesion, O'Connor and Fish (1998) the CSOS showed no significant difference between the two groups.

The authors suggested further inquiry to determine any relationship between systems and student achievement. They further advised that teacher education programs increase the hours that student teachers spend in schools with college classes geared to help students reflect about their teaching.

The 1998 study by van Driel, Verloop, and de Vos, like Schempp et al. (1998) focused on the role of subject-matter knowledge. They investigated how science teachers transformed subject-matter knowledge and how they related their transformations to student understanding. The study was based on Shulman's concept of PCK (as cited in van Driel et al.).

To achieve their purpose, the researchers designed an experimental course on chemical equilibrium for students of upper-secondary education

and an in-service workshop for chemistry teachers using the experimental course in their own classes. The grounded theory approach incorporated three cycles. Each cycle involved designing, implementing, evaluating, and reflecting on the experimental course. The workshops on chemical equilibrium were designed with constructivist views supporting and facilitating participants' construction of PCK by providing practical experiences, research results, and organized interactions.

Each of the 12 study participants had an academic background in chemistry and more than five years of experience teaching chemistry in upper-secondary education. The volunteers were familiar with the topic and wished to make their educational practices more innovative.

The researchers audiotaped all workshop sessions. Participants' written responses to assignments during the sessions and an evaluative questionnaire provided additional data. Additionally, the participants collected, corrected, and submitted the written work of their students in the experimental course. Analysis of the audio recordings utilized the stepwise procedure. This procedure involved selecting fragments relevant to subject-matter knowledge or PCK. Next, van Driel and de Vos transcribed and analyzed the fragments. Triangulation was accomplished by comparing and discussing interpretations of the individual researchers. The constant comparative method was used for the comparison and analyses of the transcripts with the other sources.

This study identified the importance of thorough subject-matter knowledge coupled with teaching experience. The researchers found that teaching experience was the major source of PCK, whereas adequate subject-matter knowledge appears to be a prerequisite. Van Driel et al. (1998) offered guidelines for teacher training programs aiming at PCK development. They asserted that these programs should provide opportunities for teachers to study subject matter from a teaching perspective through topic-related workshops.

Appleton and Kindt (1999) also explored a conceptualization of the development of PCK. They focused on elementary science teachers by reporting on two related studies. The first study involved science teaching practices of beginning elementary teachers, and the second study delved into elementary teachers' understanding and use of "activities that work" in science. During the first study the researchers discovered that some elementary teachers who lack science content knowledge used "activities that work" to generate PCK which enabled them to teach science. The purpose of the second study was to determine that if PCK was a useful and valid construct, why and how have some teachers, who were not science specialists, developed science and topic PCK sufficient to enable them to teach elementary science effectively.

The first study was a small case study involving nine recent graduates' practices in teaching science. The researchers interviewed and observed these

teachers in a variety of schools and grade levels. The second study included 20 participants from most grade levels. However, the experience of the participants in the second study ranged from a few years to 20 years. The authors did not include information on the number or frequency of observations and interviews. (No information was given about the participant selection process.) Appleton and Kindt (1999) were not specific about their method of data analysis. However, they did include many quotes from teachers' interviews and descriptions of classroom activities gained through their observations.

The researchers identified the common themes of "activities that work" as (a) an activity that taught the required content, (b) the background science content was known to the teacher, (c) involved students and was fun for them to do, and (d) had a predetermined, predictable outcome. They discovered that "activities that work" took two forms with elementary science teachers: (a) activities that employed pedagogies from other subjects and (b) hands-on activities. They further contended that these activities were not usually abandoned because they worked for the teacher. These activities became a part of a teacher's permanent repertoire and were used year after year unless they failed dramatically or the teacher encountered a lack of resources necessary for the activity.

Appleton and Kindt (1999) concluded that the use of these activities contributed toward a science curriculum that was fragmented; "a series of

largely unrelated activities which probably contributed little to progressive conceptual development in students" (p. 9). They determined that the activities were more isolated experiments than investigations, limiting the inquiry process. The authors recommended that since such a repertoire of activities was a key need for beginning teachers, teacher education programs and school systems should provide a set based on a defined curriculum and implement instruction in a more structured science curriculum as support for these teachers. They noted that most PCK research has been conducted in secondary schools where more content knowledge was expected and that PCK development in the elementary context may differ.

Summary

Every study in this section found that expert teachers employed skills, possessed knowledge, and used strategies that were identifiable. Although all these researchers did not use Shulman's (1986) concept of PCK as their theoretical framework, the attributes of PCK were evident in most of the findings. While Leinhardt and Greeno (1986) focused on the cognitive processes of expert teachers (pedagogy), others (Hastie & Vlaisavljevic, 1999; Schempp et al., 1998) studied subject matter expertise of expert teachers (content). O'Connor and Fish (1998) tied expertise to experience and found flexibility, communication, and cohesion (pedagogy) in the expert teachers.

Four studies (Appleton & Kindt, 1999; Schempp et al., 1986; van Driel et al., 1998; Ward & O'Sullivan, 1998) were based on the concept of PCK and

found evidence to support the need for both pedagogical and content knowledge. All the researchers in this section either used experience as an indicator of expertise or found that experience was related to expertise. All studies in this section either used teacher experience as a part of the criteria for selecting expert teachers (O'Connor & Fish, 1998; Schempp, Manross, Tan, & Fincher, 1998; van Driel, Verloop, & de Vos, 1998) or identified teacher experience as one of the characteristics of expert teachers (Leinhardt & Greeno, 1986; van Driel et al., 1998). While the two words *expertise* and *experience* are not synonyms, experience is used as a descriptor of expertise in most of the studies included in Section 1. Findings of all these studies are summarized in Table 1.

Table 1

Identified Characteristics of Expert Teachers

Researchers	Sample Size	Identified Expert Teacher Characteristics
Leinhardt & Greeno (1986)	15	Presentation and review, shared presentation, drill, game drill, homework, guided practice, monitored practice, tutoring, tests, transitions, action agenda lessons, set patterns of identified activities, established routines, self-monitoring, and self-regulating

Table 1 (continued).

Identified Characteristics of Expert Teachers

Researchers	Sample Size	Identified Expert Teacher Characteristics
Schempp, Manross, Tan, & Fincher (1998)	10	Recognized problems in student learning, detailed planning and organization, accommodated wider range of student abilities and skills, comfortable with and enthusiastic for teaching the skill, importance of subject matter expertise
Hastie & Vlaisavljevic (1999)	9	Importance of subject matter expertise, used application and extension
O'Connor & Fish (1998)	20	More flexible, higher level of communication, responsive, adaptable
van Driel, Verloop & de Vos (1998)	12	Thorough subject-matter knowledge, teaching experience
Appleton & Kindt (1999)	29	PCK acquisition differs between secondary and elementary teachers, elementary teachers learn strategies to compensate for lack of subject-matter knowledge

In the next section, studies on the beliefs, skills, and assimilation processes of novice teachers are examined. It is the application of these assimilation processes that allows novice teachers to move into the realm of expertise.

The Novice Teacher

The novice teacher wanders the landscape of the school campus like a yearling searching for his place in the forest. His naiveté is compelling to the researcher searching for a tidbit of information to taste and ingest, testing for flavor and value. For decades, the researcher has wandered the same forests as the yearling compiling a plethora of information on the trials, successes, and failures of the novice teacher.

This section explores the research on novice teachers' beliefs, skills, knowledge, strategies, experiences, and growth. Most of the research on novice teachers consists of small case studies providing specific insight into the novices' thoughts as they complete the first year in their profession. In this section literature relevant to the experiences and development of first-year teachers were reviewed. Feiman-Nemser (2001) described the novice teacher as knowing *about* teaching, but needing to learn *how* to teach.

Brock and Grady (2001) described the characteristics of first-year teachers: "(a) changes in the definition of oneself, (b) experiences in a totally new situation, and (c) major changes in the interpersonal support network" (p. 6). In addition, this is usually novice teachers' first foray into adulthood

and it is further complicated by their responsibility for the learning environment of approximately 20 students. Williams and Williamson (1996) portrayed the first year of teaching as one of assimilation, experimentation, and continued growth. After reviewing over fifty years of research to identify problems and issues facing beginning teachers, Zepeda and Mayers (2001) posited that, "Findings of these studies indicate that these problems and issues are perennial" (p.1).

Bullough (1987) explored the problems and responses of the first-year teacher as he or she tries to fit into an institutionally prescribed role and how these responses relate to the development of expertise. The author conducted a case study on one first-year teacher as she progressed through the year. The subject, a seventh grade teacher, was part of a three-person team. Her students included 23 remedial students in the morning and 36 average students in the afternoon. The advanced students were assigned to another team member.

The researcher interviewed the novice teacher prior to the beginning of the school year to gain information about how she perceived her role as a teacher and her concerns. The next step in the study involved weekly classroom observations. At the end of the day of the observation, Bullough (1987) again interviewed the participant using stimulated recall questions. These questions arose from observed teaching actions during the classroom observation. Additionally, he asked questions which originated from analysis

of the interview transcripts, coded to identify emerging themes. The researcher also interviewed the principal and four randomly selected students at mid-year. The author did not specify which group of students these four were selected from; only that there were two boys and two girls.

Further analysis of the identified themes revealed that the subject used six different types of responses to encountered problems: (a) environmental simplification, (b) stroke seeking and withdrawal, (c) context restructuring, (d) compromise, (e) skill improvement, and (f) laughter. These responses became habituated over the period of one year. The author termed these responses as "successful coping," but warned that there was a down side (p. 234). He posited that the better she became at coping, the less likely that the environment would become more educative professionally.

Although Bullough's (1987) study included only one participant, he completed the study with several suggestions. First, he recommended fundamental changes in teacher education programs in that they should pay more attention programmatically to students' values. Second, he contended that teacher education institutions should include instruction in institutional survival skills. Third, he asserted that teacher education should extend into the first year of teaching providing time for beginning teachers to reflect on their practice. Finally, Bullough's study indicated that beginning teachers would benefit if school systems structure their first year to include fewer

preparations, mentoring programs, and linkage with groups of teachers formed explicitly to study practice.

Brickhouse and Bodner (1992) also studied one novice teacher: a second-year middle school science teacher. They explored his beliefs about science and science teaching to determine how these beliefs influenced, or failed to influence, classroom instruction. The research centered on (a) what the teacher believed he should be doing to teach science effectively, (b) what instruction was actually carried out, and (c) what constraints hindered instruction consistent with his beliefs. This case study used audiotaped classroom observations, formal and informal interviews, field notes, and classroom records (tests, quizzes, and worksheets) for data collection over a seven-month period. The researchers used purposeful selection in choosing a relatively inexperienced science teacher.

Data collection consisted of four formal, semi-structured interviews with prepared open-ended questions. The first of these four interviews generated hypotheses which were tested by subsequent interviews. The other three interviews focused on methods of instruction and specific actions the teacher had taken in the classroom and were compared with the teacher's answers in the initial interview. The classroom observations totaled 36 hours over the seven months and were unannounced two or three times a week. All interviews and observations were converted to fieldnotes which included teacher movements as well as dialogue.

Brickhouse and Bodner were not specific about their analysis technique. However, the themes they identify as conflicting with the subject's ideas about science instruction were classroom constraints and institutional constraints. This indicated the identification of themes; implying the use of the constant comparative method.

When Brickhouse and Bodner (1992) analyzed the science teacher's beliefs they found that he believed that "scientists are curious, creative, and motivated purely by a desire to understand the natural world" (p. 475). Their study of his actual instruction found students who were encouraged to "stick to the definition given in class – not to stray from the road map." There was no encouragement for thinking beyond where they currently were (p. 477).

When the researchers investigated the constraints which hindered the teacher's instruction they found student constraints and institutional constraints. Students' concerns centered on their grades rather than on course content. The institution provided little support with mentoring, supplies, or text autonomy. In general, Brickhouse and Bodner (1992) found that most of the teacher's learning occurred over a period of time in the school in which the teacher taught. "We must therefore examine the learning of teachers in schools to fully understand why they teach as they do" (p. 482). They recommended longitudinal research which monitored change in teachers' beliefs and actions. They also recommended an examination of how teachers' experiences influenced the changes they make in instruction.

Rust's (1994) study followed Brickhouse and Bodner's (1992) two recommendations for research on change in teachers' beliefs and actions and how their experiences influenced their instructional changes. His comparative case study focused on the teachers' first year of practice with the dual purposes of discerning the congruence between their espoused beliefs about teaching and their classroom actions and determining whether these beliefs changed during their first three years. This case study of two beginning teachers comprised a small part of a much larger longitudinal study involving 41 participants. The researcher used a beliefs questionnaire, dialogue journals, and interviews to study the evolution of beliefs and teaching as these beginning teachers progressed through their first three years.

The participants completed the beliefs questionnaires at the beginning of each new school year. The questionnaires focused both on the participants' reasons for entering the field and their understandings of the "nature of teaching, learning, schooling, and the purposes of education" (p. 206). The monthly dialogue journals provided insight into the participants' thinking about their work. Rust (1994) interviewed the participants periodically for immediate personal exchange of ideas. Rust was not specific as to his method of data analysis, writing only that, "the narratives that follow are derived from their journals and responses to the beliefs questionnaire, conversations with me, and my observations in their classrooms" (p. 207).

Analysis of early questionnaires, journals, and interviews of these two participants revealed they shared images of the teacher as a facilitator, role model, and motivator. The participants believed teaching should be interactive and designed to foster student independence. They shared a belief in shared leadership by all stakeholders, but offered few comments on student-teacher and home-school relationships. The two case study participants shared these beliefs with the majority of the participants in the larger study.

As the participants progressed in their early teaching experiences, their responses illustrated belief changes. The ideas about shared leadership disappeared. Comments about home-school relationships surfaced as they blamed students' parents for their own difficulty in getting students to be active learners. Discipline and teacher control assumed major importance in questionnaires, journals, and interviews. Both were overwhelmed with responsibilities and perceived that they had little support from the administration. Both blamed the university pre-service program for not having prepared them for classroom management, discipline, and burden of paperwork. They began behaving in ways which were inconsistent with their beliefs about being a good teacher.

Rust (1994) posited that universities prepare students for "front stage behaviors" or observable teaching behaviors. They do not coach students in the "backstage behaviors," the hours of planning, networking or support

systems, and the subtleties of classroom management (p. 220). Like Brickhouse and Bodner (1992), he found that teachers learn to teach within the context of the school in which they teach. He contended that university pre-service programs must address the critical issues of beliefs, change, and leadership. Additionally, he recommended that universities and school systems provide pre-service and in-service programs which would support new teachers in their first critical years.

Munby and Russell (1994) explored the development of professional knowledge by personally revisiting the experience of learning to teach. Their ethnography involved immersing themselves into the day-to-day learning of a group of student teachers. The researchers assumed the roles of teachers, observers, and interviewers with 19 students during 20 weeks of practice teaching in physics, interspersed with campus classes. The campus classes entailed the students observing Russell as he taught a 12th-grade physics class.

Munby and Russell (1994) collected and analyzed data on the students' understandings about the process of learning to teach organizing responses under four themes: (a) expectations about learning to teach, (b) observation skills, (c) credibility of a professor who teaches every day, and (d) overall perspectives on teacher education. In addition, they analyzed free-response questionnaires completed by the students. These questionnaires addressed strengths, weaknesses, and suggestions of the model teacher. The authors did

not specify their method of data analysis, only that observation field notes and interviews were analyzed.

Findings from the study included the variety of student beliefs and the strength in which they held these beliefs. Students were either discouraged by the lack of specific information on how to teach or were bewildered by their classmates' expectation of this information. The researchers used the construct of *authority of experience* because of their contention that students do not master learning from experience during pre-service programs in a way that would give them the knowledge-in-action necessary. Munby and Russell (1994) contended that although beginning teachers were poised to move from the authority of their teachers into their own authority, this readiness was hampered by the continued authority of cooperating teachers and university representatives.

While this finding seemed to contradict the findings of Brickhouse and Bodner (1992) and Rust (1994) on continued support from university and school system, there were similarities. Munby and Russell found the beliefs systems about teaching, formed over thousands of hours in a classroom, were "acquired early and persistently reinforced" (p. 92). They suggested that universities should explore authority in teacher education to prepare pre-service teachers to move from the stance of being a student and subject-to-authority to that of taking charge as teachers and moving into positions of authority.

Schmidt and Knowles' (1995) research differed from previous studies in that they begin at the end; novice teacher failure. Their study followed the unsuccessful experiences of four novice teachers. Their study focused on the relationship between women's conceptions of knowing and learning and the work of novice teachers. This interpretive case study used extensive observations, weekly or monthly interviews, semi-structured reflective journals, and individual stories to collect data.

Four beginning teachers with differing preparation backgrounds comprised the participation group. Two were student teachers in competency-based teacher preparation programs which emphasized the mastery of specific skills for teaching music. They were both assigned to work at three different fifth- through 12th-grade schools with three cooperating teachers. The third was from a graduate teacher preparation program which provided integrated theoretical study and practical experiences. She was assigned to teach her worst two subjects, science and history. The fourth was a college graduate in history and social science who accepted a position teaching Spanish. All four had been successful students, not necessarily by learning useful information, but by being obedient and compliant in school.

Schmidt and Knowles (1995) analyzed the data in light of four factors: (a) participants' personal histories, (b) their understandings of themselves as teachers, (c) instructional problems experienced, and (d) contexts of their beginning teaching experiences. The researchers found that the perceived

"failures" of each teacher resulted, not from any single factor, or event, but from an accumulated effect of events and experiences which dated back to their childhood school experiences. All four possessed personality traits such as unassertiveness, compliance, and shyness. This contributed to their lack of success in the classroom.

Their findings reinforced those of Munby and Russell (1994) on early-formed beliefs systems and the need for universities and school systems to provide continued support. The findings of Schmidt and Knowles suggested consideration of more connected, collaborative styles of supervision or mentoring; helping novice teachers validate and give meaning to their own experiences. One significant limitation of this study involved the placement of the participants. Schmidt and Knowles studied novice teachers who accepted positions outside their areas of training. The first year of a teaching career presents difficulties to those who are well-prepared for their assignment. The difficulties encountered by novices teaching outside their field of study are even greater. Not only must they face the management and organizational issues of any first year teacher, they must also acclimate themselves to a content area unfamiliar to them. The study would have been strengthened if the researchers had selected teachers who were practicing in their areas of study.

In a study similar to Bullough (1987), Ward and O'Sullivan (1998) studied the changes in the pedagogy and content of one novice teacher as a

function of the acquisition of expertise. However, they expanded the time frame from one year to six years. They framed their research on Berliner's theory of expertise development (as cited in Ward & O'Sullivan, 1998). Berliner used experience as a key variable in the five stages of teacher development he delineated.

The researchers used a case study design to record the changes in a physical education teacher from year two of his career to year six. They viewed the teacher's pedagogy and content as a function of his experience a microanalytically. A comparison between years two and six provided a view of slightly more than half the time researchers considered necessary for expertise development. Also, the fifth year was a point at which Rosenholtz found that approximately 30% of new teachers have left the profession (as cited in Ward & O'Sullivan, 1998). (No information was provided on the selection of the subject of the study, a teacher in a lower-middle class suburb of a large metropolitan city.)

Ward and O'Sullivan (1998) used direct observations and interviews as their main sources of data. During both years two and six the direct observations and semistructured interviews focused on a basketball unit of study and a gymnastics unit. At least half the lessons were observed and videotaped during each study year. The second author conducted the interviews following each unit, utilizing the interview guide approach which entailed deriving questions from the videotapes and a priori concerns which

surfaced during the interview. The researchers then transcribed and analyzed the videotapes of the observations and interview audiotapes triangulating these with field notes and descriptive data. The teacher reviewed the resulting manuscript for approval.

The researchers organized their findings inductively into three themes: (a) pedagogical reductionism, (b) typicality, and (c) isolation. The teacher demonstrated *pedagogical reductionism* by reducing his pedagogical options to a one-size-fits-all approach to teaching with little variation. In the area of *typicality* the teacher had changed his view of what was typical, believing his students were different and less capable. The teacher's *isolation* provided few opportunities to interact with other physical education teachers with which to compare practices. The authors contended that the teacher's gymnasium world was part of a larger world including the school, the district, and the political structure. They called for more studies on the forces that drive the character of the instruction and the uniqueness each setting imposes on its players. Like Bullough (1987) they suggested that teacher preparation should extend into the novice teacher's first year of teaching.

While the other researchers on novice teachers examined general concepts, such as PCK and TPK, Ralph (1999) focused on specific instructional skills. He investigated the extent of the development of teacher-interns' oral-questioning skills. This quantitative study, which was part of a larger study, relied on a three-source data collection process after participant

training sessions. The training involved four full-day training sessions, over a four-month period, on selected instructional skills, including oral questions. The training was based on the Ralph's Contextual Supervision model (as cited in Ralph, 1999). This model involved the interaction of novice teacher and classroom cooperating teachers (CCT) with the response and guidance of the CCT depending on the competence of the novice.

In the first phase of data collection, all interns analyzed their own questioning skills used in two audiotaped lessons. The second part involved the researcher observing four lessons and sharing observation notes with the participants. The third source of data included four printed surveys of participants' responses on the effectiveness of their oral-questioning skills. Participants for the study were nine cohorts of interns and their classroom cooperating teachers (CCT) with whom the researcher worked during the years from 1994-1999 (N= 95).

Ralph (1999) collated the paired responses of each intern and CCT to produce an overall picture of the trends and patterns. The researcher then calculated the mean and standard deviation for the responses to each of the seven items included in the five-position, Likert-type scale surveys. He chose the mean and standard deviation because these were the most dependable measures of central tendency and variability. Each participant responded to: (a) value of questioning, (b) extent that questioning was used, (c) degree of student thinking required, (d) degree of wait-time provided, (e) degree of

variety of levels used, (f) degree of clarity and conciseness, and (g) degree of distribution of questions to groups.

One finding from the shared videotape analyses, observation fieldnotes, and surveys indicated that interns saw themselves as improving in all seven of their questioning competencies over the six-week period. Interns and CCTs agreed on this assessment. A second finding was that the range of variation of both subgroups' ratings decreased over time, suggesting that participants' perceptions of the interns' oral-questioning effectiveness moved toward greater congruency as the research period progressed. A third finding was that both interns and CCTs ranked the interns' performance in questioning skills in the same hierarchal order. Both ranked valuing oral questioning as the highest area of performance and higher-order learner thinking as the lowest.

Ralph (1999) determined that neophyte teachers develop instructional skills under the guidance of experienced practitioners. Brickhouse and Bodner (1992) recommended an examination of how teachers' experiences influence the changes they make in instruction. Ralph suggested one influencer of change was the guidance of experienced practitioners.

Stanulis, Campbell, and Hicks (2002) investigated the sources of knowledge a novice actually uses as she develops her teacher identity and learns to teach. The action research case study focused specifically on one novice's own questions of how she was finding her way during the induction

years. Their questions explored both her epistemology and her practice to further develop her reflective capabilities as a new teacher.

The researchers employed four teaching observations followed by loosely structured interviews and journals written by the subject. Throughout the data collection they discussed and read transcripts of previous sessions to plan further data collection. Stanulis et al. (2002) then individually analyzed the interviews and observation notes identifying themes. The subject was actively involved in the theme identification, leading the conversations, and describing her perspective on the themes.

Like Ward and O'Sullivan (1998), Stanulis et al. (2002) found isolation in the classroom and isolation from the university made the adjustment to school culture difficult. They also identified a lack of support and scaffolding from the university. One of the factors which made the transition from student to teacher smoother was the mentoring of the community of colleagues in her new environment.

"Learning by practice" is a recurring theme in the findings of several researchers (Brickhouse & Bodner, 1992; Bullough, 1987; Rust, 1994, Schmidt & Knowles, 1995). Closely related to this concept is that of risk-taking. The willingness of novice teachers to take risks by experimenting with new strategies and behaviors is an essential tenet of their ability to grow and find their identity as teachers.

Gwynn-Pauquette, & Tochon (2003) explored the idea of risk-taking by studying the reflections of 14 preservice teachers preparing for English as a second language (ESL) or social science teaching in high school. They determined what encouraged the participants to experiment with innovative practice, notably cooperative learning, during their student teaching period: what they perceived as tolerable and intolerable risk factors, and what encouraged them to continue trying the new approach.

The researchers recorded and analyzed planning and post-observation conversations with the participants. The analysis revealed that willingness to take risks depended primarily on the reactions of their students and the support they received in their teaching situation. This need for support is reflective of the findings of Rust (1994), Schmidt & Knowles (1995), and Stanulis, Campbell & Hicks (2001). Although the participants in this study were not yet at the novice stage, they faced many of the same challenges.

Lasky (in press) studied the influences that shaped professional identity. The researcher investigated early identity influencers and how the participants' professional identify affects their work with students. Surveys and interviews were the primary data sources from an urban comprehensive high school in Canada. Lasky found that how teachers were taught shaped their teaching and beliefs. These beliefs were challenged by the atmosphere in which they worked. Contributing to these challenges were the larger social and political systems which influenced core aspects of teacher identity.

Of importance to this study, were Lasky's (in press) findings on professional vulnerability. There is an emotional aspect of teaching and learning that is dependent on connections between teachers, students, and the emotional world which enables students to construct meaning, make sense of relationships, and translate their learning to the everyday world. This aspect is strengthened by teachers' willingness to be vulnerable with their students as they aid in socio-emotional development. To the teachers in Lasky's study, vulnerability and professional risk-taking were inseparable in effective teaching.

Reio (in press) conceptualized the findings of Gwyn-Pauquette & Tochon (2003) and Lasky (in press) on identity, emotional experience, and risk-taking in his conceptual model which also takes into account reform and teacher background variables. He contended that reforms and changes at play in today's educational arena have an ultimate effect on teacher learning and development. Reio's model (Figure 2) reflects many of the variables identified by other researchers in the area of effective teaching.

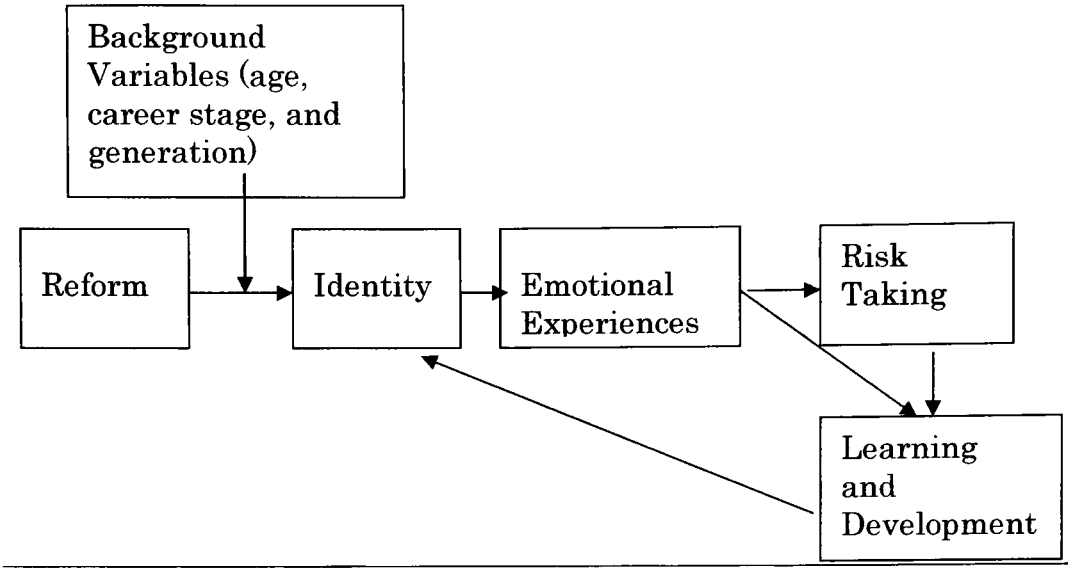


Figure 2. Reio's conceptual model of the influence of reform on teacher identity, emotions, risk-taking, and learning

From "Emotions as a lens to explore teacher identity and change: A commentary." by T. G. Reio Jr., (in press). *Teaching and Teaching Education.*

Summary

Findings from studies in this section illustrated the immensity of the changes experienced by first-year teachers. Their beliefs systems formed over nearly 20 years in a classroom as students were challenged by the constraints of their new context (Brickhouse & Bodner, 1992; Munby & Russell, 1994; Schmidt & Knowles, 1995).

Novice teachers faced the full responsibility for the academic progress, emotional security, and safety of a large group of children and were overwhelmed with responsibilities and perceived that they had little support

from the administration (Rust, 1994). Hampered by isolation (Stanulis, Campbell, & Hicks, 2002; Ward & O'Sullivan, 1998) they were not familiar with the rules imposed by the educational institution (Brickhouse & Bodner, 1992; Bullough, 1987; and Schmidt & Knowles, 1995). Novices were not confident of their abilities and were so constrained by time that there was little opportunity to seek assistance (Ralph, 1999). Most importantly, their beliefs about student learning were choked by the reality of survival (Rust, 1994) the first stage in Berliner's (1988) concept of teacher development.

The researchers, generally, suggested a need for university pre-service programs which better prepared the beginner, combined with connected, collaborative induction programs at the system level, to support the novices during their first critical years. Findings of all these studies are summarized in Table 2.

Table 2

Identified Characteristics of Novice Teachers

Researchers	Sample size	Identified Novice Teacher Characteristics
Bullough (1987)	1	Successful coping, learning by practice
Brickhouse & Bodner (1992)	1	Beliefs conflicts, learning by practice, student and institutional constraints

Table 2 (continued).

Identified Characteristics of Novice Teachers

Researchers	Sample size	Identified Novice Teacher Characteristics
Rust (1994)	2	Beliefs conflicts, overwhelmed with responsibilities, little institutional support, learning by practice
Munby & Russell (1994)	19	Beliefs conflicts, change in authority
Schmidt & Knowles (1995)	4	Beliefs conflicts, little institutional support, learning by practice
Ward & O'Sullivan (1998)	1	Isolation, pedagogical reductionism, typicality
Ralph (1999)	9	Benefit from practitioner guidance, self-perceived growth
Stanulis, Campbell, & Hicks (2001)	1	Adjusting to school culture, lack of support from university, isolation, mentor value

This section identified the behaviors learned, overwhelming responsibilities, and changes experienced in the first year of teaching. It also included a discussion of variables that influenced teacher growth and

maturation. In the next section research on specific differences between novice and expert teachers will be reviewed.

Comparing Novice to Expert

Research comparing novice and expert teachers is, with few exceptions, qualitative in nature. While the specific dimensions of each study are not identical, they all address one or more of the myriad differences which differentiate one group from the other. Researchers seek to characterize the expert teacher and assess the novice teacher's progress toward this ideal. The majority of the research comparing these two groups of professionals focused on their mental deliberations prior to, during, and after the teaching process.

One of the notable concepts which surface in most of the studies reviewed is what Shulman (1987) terms *wisdom of practice.* Without experience, novice teachers do not possess this wisdom. According to the research that follows, experience plays a key role in the development of an expert teacher.

Darling-Hammond (1995) contended that novice teachers were still trying to master a wide range of skills including motivating students, assessing progress, meeting the multiplicity of needs in a large population of students, and managing student behavior. She termed novices to be less effective than their more experienced counterparts related to the aforementioned skills.

Borko and Livingston (1989) investigated the differences in the thinking and actions of expert and novice teachers by comparing their planning, teaching, and reflections. This ethnographic study used classroom observations and interviews supplemented by photocopies of planning documents (texts, written plans, and content notes). The researchers used two theoretical frameworks for their study: (a) Leinhardt and Greeno's characterization of teaching as a complex cognitive skill determined, in part, by the nature of teacher's knowledge system and (b) Yinger's framework of teaching as an improvisational performance.

The researchers observed math instruction daily for one week for approximately one hour for both the novice and expert participants. Borko and Livingston (1989) conducted pre-observation interviews with each participant focusing on instructional planning and the nature of the lesson. They compiled condensed fieldnotes guided by a set of general questions about instructional activities, routines, and strategies. They expanded the fieldnotes immediately after the lesson, aided by audiotapes of the lesson. Post-observation interviews examined participants' reflections regarding prominent features of the lesson, unexpected occurrences, and changes from the lesson plan.

Borko and Livingston (1989) used Ethnograph, an automated coding, search, and retrieval to code and sort the data from the interviews and observations. In the first step, they identified categories of thinking and

actions. Next, they compiled a list of statements, activities, and behaviors and placed them in the correct category. In the third step, the researchers prepared a case description for each participant from the data. Finally, they examined the case descriptions for patterns of similarities and differences in participants' thinking and actions. Participants in this study included three student teachers and their cooperating teachers (N= 6). The researchers chose the three student teachers because each had a strong background in mathematics and the reputation method was used to select three expert teachers as their cooperating teachers.

In the area of planning Borko and Livingston (1989) found expert teachers' plans, while often not written, typically included a general sequence of lesson components and content. Similarly, novice teachers' *planning* included a mental plan or agenda with few details. In the areas of *interactive teaching*, the expert teachers demonstrated skill in keeping the lesson on track and accomplishing their objectives while allowing questions and comments. In addition they achieved a balance between content-centered and student-centered instruction. Conversely, the novice teachers were not as successful in implementing their plans. All three experienced difficulty when students' questions or comments led them to make explanations for which they were not prepared. This implied their plans were not thorough with little thought of questions that might arise from the students.

When the researchers examined the area of *postlesson reflections*, they found that expert postlesson reflections demonstrated conciseness and focus. Their reflections addressed student understanding. Discussing only events that had an impact on goal accomplishment, they did not mention student behavior. The researchers explained the narrow focus of the experts as a function of their ability to draw upon their well-developed systems of cognitive schemata. They attended to and processed information only if it was meaningful to their agenda.

Experts worked from a mental script and improvised their plans quickly, generating examples and drawing connections to students' prior knowledge. In contrast, the novice teachers' reflections related to the events of the day, student participation, and the behavior of the students. There was no consistent focus to their reflections. They experienced difficulty in making smooth transitions back to the lesson after student questions. Students' requests for unanticipated explanations presented problems. As a whole, unplanned explanations provided obstacles to the smooth completion of the lesson for the three novices.

Borko and Livingston's (1989) analysis suggested that "novices may not have the necessary knowledge and skills to adopt the complex patterns of teaching activities displayed by experts in the classroom" (p. 492). From the cognitive perspective, the fully developed schemata of the expert teachers contrasted with those of the novice teachers who were still in the process of

developing these schemata. The authors identified a need for further research into the development of pedagogical expertise.

In a similar study, Westerman (1991) studied the nature of expertise in teaching by comparing the thinking of expert and novice teachers during the three stages of decision making: preactive or planning, interactive or teaching, and postactive evaluating and reflecting. The theoretical basis for this grounded theory study was previous research on the cognitive processes of teachers.

Westerman worked with two groups of teachers: five novice teachers and five experts who agreed to act as cooperating teachers for the novice student teachers. The school principal and university personnel recommended the experts from a group of volunteers. Each had over five years of teaching experience in the elementary grades. All were currently teaching in a middle-class suburb of Washington, D C. The novice teachers were five undergraduate student teachers in their senior year of college. Each participant taught two lessons for the study. Subjects taught included language arts, mathematics, social studies, and spelling in grades one through six.

For each of the two lessons per teacher, data were collected in four phases. In phase one, each teacher was interviewed before teaching the lesson. The interview focused on her decision-making process in planning for the lesson. In the second phase, the lesson was videotaped and the teacher

participated in a stimulated recall interview using the video. As the teacher viewed the video, she stopped it each time she remembered making a decision and explained the thinking for the decision. The third phase occurred immediately after the stimulated recall interview when the teacher was questioned to elicit evaluation and reflection on the lesson. The fourth phase was conducted several months later when the videotape was used again, this time without sound. To capture decision making not included in the stimulated recall, each teacher was asked to talk continuously during the tape and explain the entire process.

The researcher then analyzed the 20 protocols for patterns of similarities and differences among the three stages of decision making and between the two cohorts of teachers. She identified categories and coded the transcripts. A second round of comparison and category refinement followed this, identifying additional categories and substantiating the preliminary hypotheses. This cyclical pattern continued through several phases of analysis resulting in the generation of theoretical propositions and model creation.

The most notable differences between novice and expert teachers identified by the researcher were (a) integration of knowledge, (b) student behavior, and (c) interaction among the three stages of decision making. In the area of *integration of knowledge,* Westerman (1991) found this difference between the two groups early in the analysis. This finding was important

because this skill allowed teachers to connect new learning to prior knowledge.

In this study, novice teachers and experts thought about and reacted to *student behavior* in different ways. Experts acted to prevent inappropriate behavior and addressed occurrences with voice, body language, and well-practiced skills. Novice teachers tended to ignore such behaviors until they were disruptive, at which time they stopped the class to attend to the disruption. Expert teachers saw *interactions among the three stages of decision making* and used these interactions to reach their goals. The researcher found that the experts used a mental image of their lessons to guide them through their instruction. Experts planned for possible student questions, fielded the questions appropriately, and reflected on their teaching in terms of student learning. Novice teacher decision making processes were more linear in nature. Novices planned the lesson, taught it, and evaluated it with little connection between the three dimensions.

Westerman (1991) created models to visually depict her theory of decision-making. This study and the resulting theory contributed to the field of evidence that novice teachers' thinking processes qualitatively differ from those of experts. She concluded that the major developmental difference between the two groups was their use of pedagogical content knowledge.

Sabers, Cushing, and Berliner (1991) identified differences in pedagogical knowledge among experienced and less experienced teachers as a

part of a larger research project. The researchers concentrated their work on teachers' understanding of classroom events characterized by simultaneity, multidimensionality, and immediacy. *Simultaneity* describes the need for teachers to continually monitor a large number of events, often at the same time. *Multidimensionality* refers to the large quantity of events and tasks in classrooms. *Immediacy* refers to the rapid pace of events and decisions.

The researchers used participant responses to a videotaped typical lesson for data collection. Three videocameras and monitors provided recorded views of the three characteristics of the study. Participants viewed three monitors focused on three different classroom aspects simultaneously. For each experimental session, a single participant viewed all three video tapes at the same time. After the viewing, the participants described observed instructional techniques or strategies and observed management techniques. Next, the participants viewed the videos again doing a talk-aloud describing what concerned and pleased them.

A third segment of the data collection involved the participant responses to nine questions concerning routines, content, motivation, learning environment, student attitudes, teacher expectations, teacher roles, critical thinking skills, and teacher-student relationships. Finally, the participants responded with "yes" or "no" to questions about specific details they had seen on each screen.

Participants included seven experts, four advanced beginners (first – year teachers), and five novices (individuals who were employed in other areas, but were interested in entering the teaching field). The researchers used a stringent elimination process to eliminate 44 expert teachers from the original field of 51. The selection process for novice and advanced beginners also eliminated several from the initial pool of participants to guarantee participants who were strong candidates for success in the field

Sabers et al. (1991) reached three major conclusions:

1. Experts were able to monitor, understand, and interpret events in more detail and with more insight than either novices or advanced beginners.
2. Experts, advanced beginners, and novices differed in how they attended to the multidimensional nature of the classroom.
3. Experts, advanced beginners, and novices performed similarly on tasks requiring judgments of content selection and on talks requiring memory for non-meaningful details.

Because the educational backgrounds of the experts and advanced beginners were similar, the researchers concluded that "experience is a critical factor in the development of competency in this kind of task" (p. 84). Their findings reflected the findings of Borko and Livingston (1989) in the area of expert and novice teacher interpretation of classroom life. In addition, they concluded that successful performance "is not likely to occur for

novices . . . for some time after they start teaching" (p.85). This too was in agreement with the conclusions of Borko and Livingston (1989). Data from their study had implications for continued training and support for beginning teachers. Sabers et al. (1991) suggested using expert teachers to structure experiences for novice teachers which would facilitate the development of expertise.

Needels (1991) studied whether experienced and novice elementary school teachers differed in terms of their perception and assessment of tasks related to teaching. The study contributed to the understanding of the growth in teacher thinking – from novice to experienced and the usefulness of classroom observations in teacher training programs. Their study analyzed teachers' assessment of a videotaped exemplary lesson taught by an outstanding teacher. The 51 participants included 18 student teachers, 14 beginning teachers, and 19 experienced teachers. Members of the purposefully chosen sample all had experience teaching inner-city primary grades.

Participants responded to the videotaped teacher's classroom management, teacher-student interactions, and the relevance of the lesson to the students' existing knowledge. Needels (1991) based the analysis on participant quality of writing, linguistic features, and the topics discussed. After the researcher had read and coded each response, a second reader read

and coded a random sample of 17 responses. Coding results indicated between 77% and 96% agreement.

Needels (1991) found that experienced teachers wrote the greatest number of words and expressed slightly more thoughts about lesson management. Their responses to teacher-student interaction were significantly longer than those of the student teachers. Additionally, the researcher found that a larger percentage of the experienced teachers' responses contained linguistic cues concerning contrast, change, and temporal or logical sequence. The findings suggested that experienced teachers chose topics such as the content of the interactions while the less experienced teachers discussed more frequently the affective climate and the enthusiasm of the teacher.

Needel's (1991) findings closely resembled those of Borko and Livingston (1989) who found that more experienced teachers' reflections focused on the accomplishment of lesson goals and student understanding, while novice teachers were more concerned with their own effectiveness.

Needels connected expertise and experience by stating that "Few differences were found between the student teachers and first-year teachers. These results suggested that acquiring an understanding of the complexities of classroom teaching requires perhaps several years" (p. 278). The authors speculated that beginning teachers might lack the experience and knowledge of classroom events and complexities to fully benefit from observing

classroom teachers. This speculation has significant relevance to current teacher education programs which place emphasis on classroom observations. If beginning teachers lack the experience and knowledge to benefit from classroom observations, it follows that pre-service teachers would garner even less from these observations.

Tochon and Munby (1993) investigated differences between novice and expert teachers, specifically with how their thinking reflected differences in their perception of time. The researchers used two techniques to gather information from the participants. First, they conducted a semi-structured 24-question interview with each participant covering planning, instruction, and classroom events. Next, they asked each participant to simulate planning the content of a course by talking and thinking aloud. Tochon and Munby recorded and transcribed participant responses to the interview questions and simulation exercise verbatim. They transferred any segments of text containing allusions to time to separate protocols for analysis. Participants included 23 novice and 23 expert junior high school language arts teachers (N= 46).

Tochon and Munby (1993) analyzed the data both qualitatively and non-parametrically. They coded all segments of text pertaining to time thematically and used a Chi-square cluster analysis to map the codes and the subjects. Additionally, they performed quantitative cluster analysis.

Findings from this study suggested that time management seemed more problematic for novices than for experts. The researchers concluded that experts appeared to have gained control over time, paired with a sense of flexibility and adapting to the unexpected. Novice teachers had a tendency to function with frequent, short-term planning modifications. In contrast, the experts' planning appeared more like an "open field of possibilities with which they play in synchrony and improvise" (p. 215). This finding on improvisation reflected the findings of Borko and Livingston (1989), who determined that novice teachers were not as successful in implementing their plans when faced with student interruptions and questions.

Tochon and Munby (1993) further concluded that knowing-in-action which came from learning-from-experience included a flexible time epistemology that allowed expert teachers to take different pedagogical paths at any time, modifying and making the teaching more responsive to the context. They found novice teachers had not yet developed this ability. Tochon and Munby's findings supported those of Pinnegar and Carter (as cited in Munby, 1993), who contended that time processing and reflection on time defined an important dimension of expertise. This finding was similar to the *authority of experience* investigated a year later by Munby and Russell (1994).

Cleary and Groer (1994) studied characteristics of the interactive or inflight decisions (teacher decisions made during instruction) expert and

novice teachers made as they continuously processed information during classroom instruction. They contended that teachers based these decisions that keep the class moving by monitoring various indicators which were not uniform or precise. The study sought to determine if these two groups of teachers made inflight decisions on the same basis.

This study used the stimulated recall interview for data collection. First each lesson was videotaped. Immediately following the lesson, the researcher and participant moved to a private viewing area and viewed the tape. While viewing the tape, whenever the teachers could recall what they were thinking at any time during the taped lesson, they stopped the tape and participated in a structured interview with a series of questions to encourage recall of thoughts and concerns. These interviews were audiotaped and the responses served as data for the study.

The study included 10 student health teachers and nine of their cooperating expert teachers (N= 19). Participants in this purposeful sample were chosen by the reputation method from university personnel and building principals. Videotaping and interviews for novice teachers occurred during weeks 8 and 16 of their student teaching experience. Expert teachers participated at two random points during the school year.

Cleary and Groer (1994) analyzed the audiotaped data with tallies. Whenever a teacher mentioned a particular type of concern (pupil, content, procedure, time, or materials), it was recorded on a coding sheet. Multiple

mentions earned multiple tallies. The researchers used *t*-tests and compared average total number of concerns (tallies) between expert and novice teachers. Additional *t*-tests examined sub-categories within each larger category (learning, attitude, behavior, tasks, etc.). Each category was defined and operationalized. To ensure coding reliability of the audiotaped interviews, one researcher coded all the data, eliminating the need for interrater reliability. The study would have been strengthened if both researchers had participated in coding the data.

Results from the study addressed inflight decisions regarding pupils, content, procedures, time, and materials. Overall, experts recalled 581 in-class decisions compared to 309 for the novice teachers. Cleary and Groer (1994) determined that expert teachers' ability to improvise on student cues was based on their rich knowledge of subject matter and classroom patterns. "Overall, expert health teachers employed a more complex conceptual map of the classroom, making more interactive or in-class decisions across a greater number of subcategories" (p. 113).

Cleary and Groer's (1994) findings were commensurate with other research findings on how novice teachers viewed the importance of content mastery, classroom behavior (attentiveness), and lesson plan details. Like Borko and Livingston (1989) and Tochon and Munby (1993), Cleary and Groer determined that the less-developed schemata of novice teachers made

it more difficult for them to make pedagogical decisions and to effect improvisational changes in the planned lesson design.

Manning and Payne (1996) investigated the differences of mental deliberations when beginning and experienced teachers were confronted with similar instructional episodes. They situated the study within the conceptual framework of expert – novice teaching and the consistent findings that novice and expert teachers differed in their pedagogical and content area knowledge (Berliner, 1988; Borko & Livingston, 1989).

The researchers selected two participants who were closely matched in gender, age, education level, and grade taught for their comparative case study of the two groups. Both were enrolled in Master's programs. The novice was in her first year of teaching and had outstanding student teaching evaluations. The expert had 15 years of teaching experience and was nominated by her principal.

Manning and Payne (1996) chose the Hormuth method for data collection (as cited in Manning & Payne) rather than retrospective self-reporting methods of interview, stimulated recall, and journal analysis. The Hormuth method is an *in situ* procedure for collecting processes as they occur naturally. This was accomplished with the use of hand-held tape recorders carried by the participants. They used self-talk during teaching episodes to capture teachers' thoughts and decision-making processes.

Together, the novice and expert teacher selected ten "common" teaching episodes which involved deliberate thought. These included both instructional and non-instructional episodes such as: (a) morning arrival before announcements, (b) discipline problems, (c) reading groups, and (d) large group instruction. Next, the teachers transcribed the tapes and selected five utterances they felt were most representative.

After intensive training, the teachers then analyzed these 50 samples of self-generated and self-selected utterances according to five classification schemes, previously developed by the authors. The five categories were: (a) directional states, (b) transactional analysis, (c) internal vs. external locus of control orientation, (d) facilitative vs. non-facilitative, and (e) self-directed vs. other-directed needs.

The expert teacher used more neutral, more adult, and less external locus of control self-talk. Additionally, the expert teacher used more facilitative and other-directed self-talk. Matthews (as cited in Manning & Payne, 1996) found that neutral self-statements were an aid to focusing, persistence, and completion. In this study the expert teacher used twice as many neutral statements as the novice. If Vygotsky (as cited in Manning and Payne) was correct that language spoken to self reflects and influences thinking and behavior, then one-half of the novice's self-talk was counterproductive and could have contributed to her teaching ineffectiveness.

Compared to the expert, the self-talk of the novice was (a) judgmental, (b) non-facilitative, and (c) self-directed. The authors suggested guidance for new teachers to examine their self-communication to put it in perspective and move on to higher levels of concern. The authors concluded that learning about teachers' mental deliberations provided a powerful insight for teacher education programs. Use of mental profiles of effective teachers might serve as cognitive and metacognitive models for novice or less effective teachers.

Allen and Casbergue (1997) focused their study on determining if novice teachers had adequate ability to accurately and thoroughly recall specific behaviors of their students and themselves during classroom instruction. The researchers framed their study in Van Manen's (as cited in Allen & Casbergue) three-level theory of reflectivity which relied on teacher recall to initiate the reflective process. This study used data collected through classroom observations and structured interviews. The researchers compiled detailed narrative field notes and used a checklist of student and teacher behaviors during one classroom instructional period (35-50 minutes in length). Allen and Casbergue used audiotapes of the observed instruction to triangulate the data for accuracy between the narrative and the checklist. A structured interview followed the observation to determine teachers' recall.

The study participants included four novice teachers, five intermediate teachers, and three expert teachers (N= 12). This sample of convenience involved novice teachers to whom the researchers had access and

intermediate and expert teachers recommended by their principals because of their readiness to share their recall. The novice teachers included teachers in the last year of their four-year undergraduate education program. Intermediate teachers had one to six and a half years of experience. Expert teachers exhibited excellent teaching skills and had a minimum of 10 years of teaching experience.

Analysis of the data involved detailed narrative field notes of classroom observations with time notations. The researchers designed a checklist to categorize and quantify both student and teacher activities. They used a time sampling procedure to record behaviors of students and teachers every five minutes. Class sessions were audiotaped and compared with the other two documentation sources to triangulate the data. Using their detailed narrative, audiotapes, and checklist the researchers compiled a sequential list of specific teacher and student behaviors.

Additionally, the authors conducted a stimulated recall with the participants immediately after the observations. Teacher responses during these interviews were compared to the list of behaviors. The researchers compared the teacher recall from the interview to the compiled sequential list of observed behaviors to determine the accuracy and thoroughness of the teacher recall. Allen and Casbergue (1997) established definitions and guidelines for both accuracy and thoroughness in quantifying differences as minimal, notable, substantial, or extreme.

Allen and Casbergue (1997) reported their findings by seven categories: (a) accuracy, (b) thoroughness, (c) focus of recall, (d) fluidity and certainty, (e) consistency, (f) general vs. specific recall, and (g) what was or was not recalled. Generally, novice, intermediate, teachers exhibited only minimal inaccuracies in their own and their students' behavior. Experts exhibited 100% accuracy. Like Borko and Livingston (1989), Allen and Casbergue found that novice teachers were not as thorough as the experts in their recall focus, recalling more of their own behaviors than their students' behaviors. Intermediate and expert teachers showed an increase in focus, over the novice teachers, on student behaviors. The findings illustrated the presence of a continuum from general recall lacking thoroughness in novice teachers to specific and thorough recall in expert teachers.

Another finding indicated that novice teachers primarily recalled neutral behaviors with differing levels of specificity. In contrast, intermediate and expert teachers recalled events more holistically with positive, neutral, and negative behaviors. The researchers noted that although the expert teachers often recalled in a more holistic or general manner, they were able to be extremely thorough when asked additional probing specific questions. This study found that teachers develop in their accurateness and thoroughness of recall as they gain experience in teaching.

These findings suggested that the experience level of teachers and their ability to recall accurately and thoroughly were strongly related. If, as

Allen and Casbergue (1997) stated "recalling specific classroom behaviors . . . has consistently been described as a first step in reflection" (p. 744), the development that occurs in recall ability of novices as they move toward expertise had significance for teachers as they learn to engage in reflective teaching.

Schempp, Tan, Manross, and Fincher (1998) identified cognitive differences between competent and novice teachers. The researchers undertook this study to test Berliner's theoretical propositions on the two major stages of pedagogical expertise – novice and competent (as cited in Schempp et al.). Data collection involved three structured interviews conducted by a team of two investigators. The first interview explored the teacher's subject matter knowledge, educational experiences, and professional career history. The second interview required teachers to plan a hypothetical instructional unit. In the final interview, the teachers described teaching a skill within the planned unit.

Five novice teachers and five competent teachers participated in the study (N= 10). The competent teachers had five or more years of teaching experience, were recommended by peers or university faculty, and had acceptable service as a cooperating or mentor teacher. The five novice teachers were student teachers nearing graduation.

Data analysis involved analyzing audiotaped transcripts of the interviews using the constant comparative method to identify trends, themes,

and categories. The researchers analyzed these data for underlying uniformities to identify differences between novice and competent teachers' knowledge. Finally, these uniformities were compared with Berliner's theory of the acquisition of teaching expertise (as cited in Schempp et al. 1998).

Schempp et al. (1998) identified three characteristic differences between novice and competent teachers: (a) perceptions of student learning difficulties, (b) conceptions of knowledge, and (c) reflective practice. The novice teachers tended to see student background as the cause of their learning difficulties while competent teachers attributed student problems to their own lesson structure and organization. Competent teachers were quicker to acknowledge their knowledge inadequacies and more willing to learn than the novice teachers who reported little use of assessment procedures in planning or monitoring their progress or that of their students. In the area of reflective practice, competent teachers recognized that students came to them with a wide range of knowledge and ability. Conversely, novice teachers tended to perceive little variation in the students.

Overall, the findings of Schempp et al. (1998) suggested that novices tended to distance themselves from the responsibilities of pedagogy, identifying environmental and societal conditions or the students themselves as the sources of any lack of classroom learning. Contrasting with this finding on novice teachers, competent teachers believed they "held the key to student success or failure, and were thus unwilling to quit on a student until all

possible options appeared exhausted" (p. 18). They were in constant search for new ideas and methods for teaching familiar subject matter, managing classrooms, and assessing student learning. These findings imply a need for school systems to provide opportunities for teachers to share ideas, knowledge, and skills.

Chen and Rovegno (2000) examined the characteristics of expert and novice teachers' constructivist-oriented teaching practices while using a movement approach in teaching elementary physical education. Chen and Rovegno employed Brooks and Brooks' (1993) list of characteristics of constructivist teaching strategies: (a) engaging students in exploratory, self-regulated, and cooperative learning activities, (b) inviting students to decide their own learning tasks and objectives, (c) asking thoughtful and open-ended questions, (d) guiding students to elaborate on initial thoughts, (e) structuring learning experiences around a "big picture", (f) organizing learning experiences relevant to students' prior knowledge, and (g) guiding students to work together productively.

This study used transcripts from two formal interviews, transcripts from 18 videotaped lessons, and coding information from the Educational Games Observation Rubric (EGOR) for the collection of data. Participants included three expert and three novice teachers (N= 6). University faculty members in Florida recommended the three expert teachers who: (a) used constructivist-oriented movement approaches to teaching elementary

physical education for more than five years, (b) published either lesson plan books or articles in professional journals, (c) designed the curriculum guides for their districts, and (d) presented workshops at state and regional conventions. The novice teachers attended Florida universities and were competent student teachers based on their ability to use Laban's movement framework to design learning tasks. They had also successfully completed training in the use of teaching strategies compatible with constructivism.

Chen and Rovegno (2000) pilot tested the 19-item-EGOR with non-participants and adapted it to meet the needs of this particular study. Ten items were used for the study. In addition, the internal consistency of the 10-item EGOR was determined by means of Cronbach's alpha reliability and item-to-total correlation on the data from the 18 videotaped lessons taught by all six teachers. The authors checked intraobserver agreement until the reliability rate ranged from 81% - 90%.

The researchers videotaped each of the six participants teaching three classes, for a total of 18 videotaped lessons. The first author then viewed the tapes and coded and rated each one using the EGOR instrument. Next, the researchers conducted two, 45-minute formal interviews with each participant. The first interview probed for teacher background and beliefs about learning and teaching. The second interview focused on gaining insight into the teachers' perspectives on the three major ideas reflected in the EGOR. Each interview was tape-recorded and transcribed for analysis.

The data analysis included both qualitative and quantitative components. Chen and Rovegno (2000) analyzed the 18 videotapes quantitatively with EGOR. They analyzed the 18 lesson transcripts and 12 interview transcripts qualitatively. One researcher read and re-read the transcripts, highlighting instances of constructivism and instances when constructivist strategies were not used correctly. That researcher then combined similar instances into categories and later into emerging themes. Next, the second researcher then reviewed all qualitative coding for confirmation or no confirmation.

Chen and Rovegno (2000) found the expert teachers' teaching practices went beyond engaging students. They facilitated and mediated learning. Novice teachers, in contrast, equated exploratory activities with constructivist teaching. Second, they found that expert teachers used metaphors, examples, or images to introduce new content. They asked questions that provided links to students' prior knowledge. Contrarily, novice teachers did not use metaphors, examples, or images. Questioning related to the task or new content. Third, expert teachers were more likely to ask students to share ideas about performance or expand ideas. Novice teachers encouraged students to share ideas but seldom requested students to expand on ideas. Chen and Rovegno's findings were closely related to those of Borko and Livingston 11 years earlier identifying teacher interactions with students as a major difference between experts and novices. This ability to improvise

and follow-up on student questions with expansions of both student and teacher ideas also coincides with the findings of Cleary and Groer (1994).

Chen and Rovegno (2000) suggested that educational institutions provide pre-service teachers with a "scaffolding" technique or supportive framework to guide them in (a) assisting students with self-regulation, (b) guiding students to use reflection and critical thinking, (c) linking new learning to prior knowledge, and (d) guiding students to work cooperatively in groups.

Summary

The literature in this section comparing novice to expert teachers reflected a general agreement on the part of the cited researchers that there is a significant difference in the areas of instructional planning, mental deliberations, strategies, and reflections (Borko & Livingston, 1989; Manning & Payne, 1996; Tochon & Munby, 1993; Westerman, 1991). Novice teacher plans contained fewer details than the experts (Borko & Livingston, 1989; Sabers et el., 1991). Novice teachers were less able to tend to the multidimensional nature of the classroom (Sabers et al., 1991; Westerman, 1991).

Novice teachers' questioning techniques, in-class decisions, and constructivism strategies lagged behind those of the experts (Chen, 2000; Cleary, 1994; Needels, 1991; Westerman, 1991). In addition, novices had not developed their recall skills to the degree that they were capable of accurate

and thorough reflections (Allen & Casbergue, 1997; Borko & Livingston, 1989; Westerman, 1991). Novice teachers, in contrast to experts, did not accept responsibility for student learning (Schempp et al. 1998; Westerman, 1991). These findings on the differences between the two groups are long-standing. With this knowledge about the disparities, what are the implications for student achievement in this time of increased accountability for student achievement? In the next section literature is reviewed that investigates the relationship between expert teachers and student achievement.

Expert teachers depend on well-organized knowledge that determines what they notice and how they solve problems. They have acquired a set of strategies that operate across all domains. They are more likely than novices to recognize meaningful patterns of information. Because of this, their actions and reactions begin at a higher place (deGroot, 1965). Expert teachers know how to tap into students' prior knowledge. They have acquired PCK (Shulman 1986; 1987) and not just content knowledge. Table 3 summarizes the finding of all these and other researchers in this sub-section.

Table 3

Comparing Novice Teachers to Expert Teachers

Researchers	Sample Size	Areas of Identified Differences
Borko & Livingston (1989)	6	Planning, interactive teaching, balance (content/student) postlesson reflections, cognitive schemata, improvisational skills, knowledge base
Westerman (1991)	10	Integration of knowledge, classroom management, decision-making, planning, improvisational skills, reflecting
Sabers, Cushing, & Berliner (1991)	16	Monitoring events, classroom management, pacing events and decisions
Needels (1991)	51	Cognizance of lesson management, sequencing, content of interactions with students

Table 3 (continued).

Comparing Novice Teachers to Expert Teachers

Researchers	Sample Size	Areas of Identified Differences
Tochon & Munby (1993)	46	Time management, flexibility, improvisational skills, planning, reflections
Cleary & Groer (1994)	19	Improvisational skills, subject matter knowledge, classroom management, planning
Manning & Payne (1996)	2	Quality of self-talk, level of mental deliberations
Allen & Casbergue (1997)	12	Accuracy and thoroughness of recall of student actions and learning
Schempp, Tan, & Manross (1998)	10	Perceptions of student learning difficulties, conceptions of knowledge, reflections
Chen & Rovegno (2000)	6	Practices (questioning, idea expansion, and constructivism)

Note. Nine of the 10 studies in this table either used experience as a selection criterion or found that experience accounted for some of the differences found

Relating Teacher Expertise to Student Achievement

Dunkin and Biddle (1974) classified four types of variables in research on teaching. *Product* variables were outcomes, such as student achievement. *Process* variables included the instructional interactions between students and teachers, such as specific instructional strategies. *Presage* variables were the properties or characteristics of teachers that they bring with them which might influence teaching, such as level of educational attainment. *Context* variables were those that can exercise direct effects on instructional outcomes, such as the size of the class.

Presage variables have been the focus of much of the early research on teaching, including Brophy and Good (1986), who identified traits of effective teachers. Researchers seemed to have left this group of *presage* variables and based their work on other presage variables; specifically professional preparation, strong subject matter knowledge, stability, and pedagogical skills (Rowan, Correnti, & Miller, 2002). This section includes studies that focus on these presage variables.

Researchers, more recently, have focused on the process variables active in every classroom. Their studies have involved specific instruction strategies, classroom management, teacher decision-making process, and professional development. In the first sub-section the researcher reviews studies relevant to the presage variable. I use the term *teacher characteristics* as the sub-section heading. In the second sub-section, studies

related to process were discussed. *Teacher Behaviors* is the title of the second sub-section.

In the following sub-sections I review the literature germane to these two groups of variables and how they relate to the first identified variable – product. *Product*, in the educational atmosphere of 2004, is student academic achievement. With increased accountability measures, punctuated by the No Child Left Behind Act (P. L. 107-110) schools and school systems are increasingly aware of the factors influencing student achievement.

Teacher Characteristics (Presage Variables)

If expert teaching can be identified, it follows the results of expert teaching can be measured. The most usual method for measuring teacher expertise is comparison of student test scores using teacher characteristics as the independent variables. What do these measurements show? Several research studies have measured the impact of teacher expertise with some consistency of outcomes. Conclusions are not always in perfect agreement on every indicator, but most agree that teacher expertise is a strong determinant of student learning. In this sub-section I review literature germane to presage variables and their relationship to student achievement. These presage variables include teaching experience, education level, certification, and teacher scores on standardized tests.

A review of the literature on teacher effect must include the work of Hanushek. For at least 21 years, economist Hanushek published an array of

education production function studies comparing resources to student achievement. These resources included teacher experience and education because, historically, teacher salaries have been tied to these two characteristics. On the whole, he found no consistent or positive relationship to school spending and student achievement (Hanushek, 1989). While Hanushek did find some positive correlation between teacher experience and student performance, he attributed that correlation to the more experienced teachers selecting teaching assignments in higher socio-economic schools (Hanushek, 1993). Like Coleman, his findings were actively debated in academic, legal, and public policy arenas, giving birth to an abundance of studies which either disputed or verified his findings.

Hedges, Laine, and Greenwald (1994) reanalyzed the studies used by Hanushek (1993) in his meta-analysis with different results. They determined that per-pupil expenditures, teacher experience, teacher salary, administrative inputs, and facilities affect achievement.

Hanushek (1996) continued the written debate by updating his sample of studies. He again found that the bulk of studies showed no significant relationship between resources and achievement.

Greenwald, Hedges, and Laine (1996) countered by assessing both the direction and magnitude of any relationship between a variety of school inputs and student achievement. Their meta-analysis of research spanning over a quarter century included studies selected by the following methods:

(a) reassessment of the studies used in Hanushek's (1986) meta analysis (as cited in Greenwald et al.), (b) search of electronic databases, (c) literature reviews, and (d) citations of sources identified by the first three methods. The researchers used 60 primary research studies which were aggregated at the school or system level and were either controlled for socio-economic differences or longitudinal in nature. Their stringent criteria for selection included six decision rules which narrowed the initial field of 2,000 abstracts to the final 60 studies.

Data analysis involved grouping variables from all the selected studies into three general categories: (a) expenditures, (b) teacher background characteristics, and (c) size. The researchers further subdivided these categories into seven subcategories or variables. For data analysis, they used two meta-analytic methods – combined significance testing and effect magnitude estimation. Greenwald et al. (1996) conducted a separate analysis using at least one of these methods for each of the seven identified independent variables: (a) per-pupil expenditure, (b) teacher salary, (c) class size, (d) school size, (e) teacher ability, (f) teacher education, and (g) teacher experience. The dependent variable in each analysis was student achievement. Because neither input nor output variables were typically measured on the same scale in all studies, the researchers used the regression coefficient. This measured the number of standard deviations of change in output which can be associated with a standard deviation in input.

Greenwald et al. (1996) found evidence of a strong and consistent positive relationship between per pupil expenditure and student achievement. They also determined that smaller schools and smaller classes exhibited a positive relationship to student achievement. Teacher ability, teacher education level, and teacher experience all correlated positively to academic achievement of students. Finally, the correlation between teacher salaries and student achievement proved large enough to have important implications for policy.

Greenwald et al. (1996) contended that their research findings only confirmed the obvious – "that money is positively related to student achievement." They asserted that more researchers should address "how money matters" by further researching the magnitude of input effect (p. 385).

This debate between researchers was indicative of the multilevel nature of school effects and the variety of variables. All four studies relied on the findings of other researchers who methodologies were as varied as their results. In 2002, Hanushek, along with Rivkin and Kain, seemed to reverse his original stance and paid tribute to the importance of teacher experience and quality (Rivkin, Hanushek & Kain, 2002).

Recent researchers in this area have conducted large-sample original studies to determine the effects of presage variables. Darling-Hammond (2000) examined the ways in which teacher quality indicators and other school inputs were related to student achievement across states. The

researcher used both qualitative and quantitative data from a 50-state survey of policies, state case study analyses, the 1993-94 Schools and Staffing Surveys (SASS), and the National Assessment of Educational Progress (NAEP). These studies provided the basis for regression analyses of school variables on student achievement scores.

Participants in the study (the SASS database) included 65,000 teachers, 13,000 school principals, and 5,600 school districts. Questionnaires from participants included data on teachers' degrees, majors, certification status, teaching assignments, average class size, salary schedules, and conditions of hiring. Additional data on each state included policies regarding teacher education and licensing, as well as state school spending data. Data from NAEP included state average achievement scores for 4^{th} and 8^{th} grade for both math and reading in 1992 and 1994 and student poverty rate.

Darling-Hammond (2000) established teacher quality (independent) variables as the percentage of teachers with full certification, a major in the subject they taught, uncertified newly-hired teachers, the percentage of teachers with master's degrees, and class size. Student achievement was the dependent variable.

The first step in the data analysis involved conducting bivariate correlations of school resource variables and student demographics with state average student test scores to examine relationships among variables and

select variables for inclusion in the multivariate equations. These analyses confirmed the following findings:

1. Student characteristics such as poverty, non-English language status, and minority status are negatively correlated with student outcomes, and usually significantly so.
2. Student characteristics are generally not significantly correlated with state per-pupil spending or with teachers' salary schedules.
3. Teacher quality characteristics such as certification status and degree in the field to be taught are very significantly and positively correlated with student outcomes.
4. Per-pupil spending (measured as current expenditures) shows a significant positive relationship with student outcomes in 4th grade reading in both years, but no relationship with student outcomes in mathematics.
5. Other school resources, such as pupil-teacher ratios, class sizes, and the proportion of all school staff who are teachers, show very weak and rarely significant relationships to student achievement when they are aggregated to the state level. (p. 28)

The most consistent highly significant predictor of reading and mathematics achievement each year was the proportion of well-qualified teachers (those with full certification and a major in the field they teach).

Darling-Hammond (2000) concluded that states interested in improving student achievement should attend to the preparation and qualification of new teachers hired to teach their students. A second conclusion was that systems should focus on the retention of these teachers. This conclusion has implications for schools systems as they build professional development and retention programs of their newly-hired teachers. Additionally, Darling-Hammond contended that systems should address improving the qualification status of teachers on staff who lacked proper qualifications.

The variables studied by Darling-Hammond (2000) and Greenwald et al. (1996) did not coincide in all cases. Both studies investigated the correlation of per-pupil expenditures with student achievement. While Darling-Hammond found no significant correlation, Greenwald et al. found a strong and consistent positive relationship between these two expenditures and student achievement. Greenwald et al. found a strong correlation between school and class size and student achievement, while Darling-Hammond found this relationship weak and rarely significant. Teacher ability, teacher education level, and teacher experience all correlated positively to academic achievement of students in both studies. Finally, the correlation between teacher salaries and student achievement proved large enough to have important implications for policy in Greenwald et al.'s study whereas Darling-Hammond found the correlation to be weak and rarely significant.

Although Greenwald et al. (1996) contended that their research findings only confirmed the obvious – "that money is positively related to student achievement" (p. 385), Darling-Hammond's (2000) only correlation between the two would be recruiting, retaining, training, and retraining a quality teaching force.

Goldhaber and Brewer (1996) systematically explored the relationship between student achievement and schooling inputs. Their research bridged the gap between teacher characteristics studies and teacher behavior studies

by examining both. The study was precipitated by earlier educational production function models used earlier by Hanushek (as cited in Goldhaber & Brewer) and teacher characteristic research by Murnane and Phillips (as cited in Goldhaber & Brewer) They broadened Murnane and Phillips' focus by adding a set of variables describing teacher behavior. Use of the *National Educational Longitudinal Study of 1988 NELS* permitted the researchers to estimate a variety of econometric models including one-way fixed and random effects models.

Goldhaber and Brewer (1996) used data drawn from the NELS. This data allowed the researchers to link students to specific class and teacher. They limited their data collection to students who completed the mathematics achievement test in the 8th and 10th grade. There were 5,149 10th-grade student participants from 638 public schools and 2,245 mathematics teachers.

Teachers' characteristics variables included teacher's degree level, years of experience, BA or MA in math, and classroom variables included class size and percentage of minority students. Teacher behavior variables included use of subgroups, effective questioning, emphasis of problem solving, and curriculum content control. The dependent variable was the 10th-grade mathematics test score.

The researchers estimated a variety of econometric models with fixed and random teacher effects and auxiliary regressions to regress estimated

teacher fixed effects on these characteristics. Traditional ordinary least squares (OLS) production function models revealed some educational resources to be significant in influencing 10th-grade mathematics scores. Auxiliary regression models showed observable teacher characteristics influenced student mathematics achievement. Like Murnane and Phillips (1981) the authors found that certain aspects of teacher behavior influenced student achievement. These included: (a) teacher feels well prepared and (b) teacher uses oral questions frequently.

The most notable finding on teacher behaviors was that scores of students with teachers who had no control over their curriculum or technique were significantly lower. Observable teacher characteristics findings included: (a) students with more experienced teachers had higher scores, (b) female teachers were associated with higher scores, (c) teachers certified in math correlated positively with higher scores, and (d) black teachers were associated with lower scores.

Goldhaber and Brewer (1996) were careful to distinguish between observable characteristics (listed previously) and unobservable characteristics, or classroom strategies and behaviors. Their results suggested that while observable variables account for a relatively small fraction of test score variation, the unobservable characteristics were important. Additionally, their results indicated that there appeared to be no correlation between observable and unobservable characteristics.

Okpala, Smith, Jones, and Ellis (2000) also examined the impact of selected educational resources and student/family demographics on fourth-grade students' reading and math achievement scores. Three key questions guided the study:

1. Is there a significant relationship between selected school characteristics and student achievement scores?
2. Is there a significant relationship between selected teachers' characteristics and student achievement scores? and,
3. Is there a significant relationship between selected student/family characteristics and student achievement scores?

For this quantitative study, the researchers obtained end-of-grade test scores for fourth-grade students from the system Board of Education Office. These scores constituted the dependent variable. Data for the independent variables included: (a) school size, (b) percent of teachers with master's degree, (c) percent of teachers with more than 10 years' teaching experience, (d) percent of students on free or reduced lunch program, (e) percent of parents with post-high school education, and (f) parental volunteer hours per 100 students. Fourth-grade students enrolled in 42 public schools in one North Carolina county participated in the study (N= 4,256).

Okpala et al. (2000) used measures of central tendency and dispersion and Pearson's Correlation Coefficient analysis to determine the significance of the selected variables on the fourth-grade students' achievement scores.

The researcher analyzed the mean and standard deviation values of both the independent and dependent variables.

Results of the study illustrated that class size and school size significantly related to fourth-grade reading achievement. However, these factors were not significant in math achievement. The percentage of teachers with a master's degree correlated positively with math scores, but was insignificant in the area of reading. The percentage of teachers with 10 years of teaching experience correlated positively with both reading and math achievement. The free or reduced percentage correlated negatively with both reading and math at a high level of significance. Post high school education demonstrated a positive correlation with reading and math achievement. The variable of parent volunteer hours was insignificant in both subjects.

Okpala et al. (2000) indicated a strong link between selected characteristics studied and student achievement. This link indicated the importance of student access to schools with experienced and competent teachers. The researcher suggested further study of other variables including teacher certification status.

While some of the independent variables Okpala et al. (2000) studied were the same as those identified by Greenwald et al. (1996) and Darling-Hammond (2000), Okpala et al. included several others with some similar findings and some different. Although the findings of these three studies contrasted with each other in several areas, they all concluded that the

quality of the teaching staff had a positive correlation with student achievement.

In 2002, Rivkin, Hanushek, and Kain responded to the request of Greenwald et al. (1996) for research on the magnitude of input effects. They conducted a study to disentangle the separate factors influencing achievement. Their study was conceptually grounded in research on educational production function. Among the questions they sought to answer, were three on teacher characteristics. First, were there significant differences among teachers in their abilities to raise achievement? Second, how important were these differences in teacher quality in the determination of student outcomes? Third, were these outcomes related to observable teacher characteristics, including teacher education and teacher experience?

The data set Rivkin et al. (2002) employed in the education production function regressions combined test score data with information on teachers and schools. The main approach of the study involved estimation of complicated fixed effects models which removed constant effects from families and schools and then focused on how variations in key factors influenced student academic achievement growth.

The researchers used data from the University of Texas Dallas (UTD) Texas Schools Project encompassing 200,000 students in over 3,000 public schools. These data included Texas Assessment of Academic Skills (TAAS) for third- through sixth-grade students from 1993 through 1995. For this study

Rivkin et al. (2002) used the data on students who remained in the same school for both fifth and sixth grade and who completed the TAAS mathematics exam in fourth, fifth, and sixth grades. This included 939 schools.

Independent variables included teaching experience and teacher educational attainment level. The dependent variable was student math achievement gains.

Student data were merged with information on teachers' experience of level of educational attainment. Matching individual students with teachers was not possible; therefore, the researchers matched students to school and grade. The study would have been strengthened if these specific sets of data could have been matched.

Rivkin et al. (2002) found support for the idea that teachers in their first two years of experience do substantially worse than more experienced teachers in 4th and 5th grade, but not in the 6th. Consistent with previous findings, the researchers determined that there was little or no evidence that post-graduate work raised the quality of teaching. The findings of Rivkin et al. were similar with those of Okpala et al. (2000) in that teacher experience significantly influenced student achievement.

The models and data set used in the study allowed the authors to draw three main conclusions: (a) teachers and schools matter importantly for student achievement, (b) there were large differences in teaching quality, and

(c) achievement gains were related to observable teacher and school characteristics, but the effects were generally small and generally most noticeable with younger students.

Rowan, Correnti, and Miller (2002) also reported on a series of analyses of a large data set from "Prospects: The Congressionally Mandated Study of Educational Opportunity." This large-scale survey compiled data on instructional processes and student achievement in a sample of United States elementary schools as a part of a government study of Title I programs.

Title I is a compensatory education program under the Elementary and Secondary Act (ESEA) of 1965. The purpose of their report was to answer questions about the size of teacher effects on student achievement. These effects included teachers' professional credentials and experience.

The researchers developed a three-level, cross-classified, random effects model to analyze data on two cohorts of students included in the *Prospects* data set. They decomposed the variance in students' growth in achievement in mathematics and reading into variance among several subsets including students within classrooms.

The large *Prospects* data set included thousands of students nationwide. The authors were not specific as to the exact numbers. Also, they did not specify if all the students included in the data set were Title I students.

Rowan et al. (2002) focused on three independent variables to measure teachers' professional background and experience. One was whether the teacher had any special certification in math or reading. The second was a measure of whether the teacher had a Bachelor's or Master's degree. The third variable was teacher experience, which served as a proxy for teachers' professional knowledge. The dependent variable in the study was students' growth in academic achievement

Rowan et al. (2002) conducted the analysis using a three-level hierarchical linear model (HLM) of students' growth in academic achievement. They developed a longitudinal data set for two cohorts of students in the *Prospects* study: students passing from grades one through three over the course of the study and students passing from grades three through six. Using these data, they estimated an explicit model of students' academic growth using the statistical methods of Bryk and Raudenbush, (as cited in Rowan et al.) and the computing software HLM/3L. Next, they estimated separate growth models for each cohort of students in reading and mathematics for a total of four growth models. The researchers measured achievement by scale scores provided by the test publisher.

In reading, neither teacher educational attainment level nor certification status showed any statistically significant effect on student achievement growth. However, teacher experience was a statistically

significant predictor of achievement growth in students. The *d*-type effect for the early grades was $d = .07$, and for later grades was $d = .15$

In math, neither teachers' educational attainment level, certification status, nor experience level showed any statistically significant effect on students' achievement growth in the early grades cohort. However, in the later grades cohort, teachers' experience level was statistically significant with a positive effect of $d = .18$.

Findings from this study suggested that "the classrooms to which students are assigned in a given year can have nontrivial effects on students' achievement growth in that year" (p. 6). Additionally, the researchers found that "in any given year, students are deflected upward or downward from their expected growth trajectory by virtue of the classrooms to which they are assigned" (p. 7). They concluded that if some students were consistently trajected upward and some were consistently trajected downward the cumulative effects of classroom placement could greatly affect academic growth.

Provasnik and Stearns (2003) explored the question of whether a single highly qualified teacher in a critical subject makes a lasting difference in the academic career of his or her students.

The researchers used the *National Education Longitudinal Study of 1988* (NELS: 88). Follow-up data were collected in 1990, 1992, 1994, and 2000. The 2000 survey, used for this study, included 12,000 8th-grade

students. Student cases that lacked 8th-grade mathematics teacher background data were dropped allowing the researchers to focus on the relationship between mathematics teacher quality and the three outcome variables or dependent variables. These were: (a) mathematics coursetaking in high school, (b) high school completion, and (c) postsecondary attainment. To operationalize these variables, the researchers used student transcript files and coded them for analysis and preparation in the three identified areas.

Provasnick and Stearns (2003) created a composite for the independent variable of teacher quality using the following survey items: (a) whether or not teachers had a graduate or Bachelor's degree in mathematics or were certified in mathematics, (b) their years of teaching experience at the elementary or secondary level, and (c) how diligent they were in keeping records on, correcting and returning, and discussing homework. Teachers were then classified as low quality, average quality, or high quality according to their responses to the survey.

The researchers used crosstabs to identify discernible patterns in the quality of a student's 8th-grade mathematics teacher and future mathematics coursetaking and educational attainment patterns. Next, they estimated the impact of teacher quality on these outcomes using ordinary least-squares (OLS) regression models. They controlled for student background characteristics and ability level along with other school factors.

Preliminary results of a simple bivariate analysis revealed that a positive correlation between the quality of 8th-grade teachers and students' likelihood of completing high level mathematics courses, completing high school, and going on to complete a bachelor's degree. However, the results of the OLS regressions indicated that the significance of the 8th-grade teacher was not as important in the area of high-level math courses completed when the data were controlled for other factors such as socio-economic status, class size, and urbanicity. Teacher quality was still a significant determinant of highest level of postsecondary education even after the regression for other factors.

Although quality teachers seemed to have no discernible impact on students' higher level math course completion, the results suggested that having a high quality 8th-grade mathematics teacher increased the students' chances of being in a high quality environment which could have an effect on future educational attainment. "The quality of the teacher may be a marker of other factors, which are more direct determinants of later educational attainment" (Provasnic & Stearns, 2003, p. 23).

In 2003, Wayne and Youngs conducted a meta-analysis of 21 studies with the objective of creating a clear interpretation of the relationship between teacher characteristics and student achievement gains. The researchers chose studies based on four criteria: (a) the data collected must address teachers' characteristics as well as standardized test scores of their

students, (b) data were collected in the United States, (c) the design must include prior achievement, and (d) the design accounts for student socioeconomic status.

They used the synthesis method to analyze the data because it required judgments about the strengths and weaknesses of individual studies and it allowed them to consider groups of studies that focused on particular teacher characteristics. Wayne and Youngs (2003) grouped the 21 studies by five teacher characteristics: (a) teachers' college ratings, (b) teacher test scores, (c) course taking and degrees, (d) certification, and (e) other characteristics.

For each characteristic, they: (a) described all relevant studies and findings, (b) rendered joint interpretations, and (c) considered implications for policy and future research. In the areas of teachers' college ratings they determined that some relationship exists between college ratings and student achievement gains. When they examined teacher test score studies the findings were divergent (five positives and two negatives). The researchers determined that the difference probably originated in the fact that the two negative studies controlled for college ratings.

Wayne and Youngs' (2003) examination of the research in the area of degrees and coursework (two studies) revealed no conclusive results. High school math showed a positive correlation, and there was no evidence for elementary schools. Studies on teacher certification (also, two in number)

demonstrated that students learned more mathematics when their teachers were certified in mathematics. Studies on other characteristics included experience and race with no conclusive evidence.

The implications from the findings in each area included a need for further research in the correlation between teacher characteristics and student achievement. Saying that studies in this area which use convincing research designs either did not exist or were not conclusive, Wayne and Youngs (2003) included no studies on teacher classroom characteristics or methods. They did, however, include the Greenwald et al. (1996) and Darling-Hammond (2000) study in their review.

McElroy and Pai (2003) also studied the impact of teacher experience and educational attainment on student performance. They used the educational production function literature as a conceptual framework for their quantitative study; this provided an opportunity to learn more about the role teachers played in students' academic success. The authors were not specific when describing their research design for this quantitative study.

The researchers disaggregated student test scores from the Texas Academic Assessment of Skills (TAAS) at the campus level using data provided by the Texas Education Agency from 1994-2001. Their sample consisted of 16,718 observations, each of which was a school in Texas in a specified year. They limited their sample to schools which had a minimum of five students in each of grades three, four, and five. Teachers at these schools

were divided into five groups according to their experience levels and pass rates were determined for each group.

The independent variables for this study were teacher experience and teacher educational attainment. The dependent variable was the percentage of students at any given school who passed the TAAS test in a given year.

McElroy and Pai (2003) determined that teacher experience showed a significant impact on pass rates. Their results lent credence to the idea that the experience of teachers was a critical input to the educational production function. They also found that the results for the effect of teacher educational attainment were mixed. The researchers suggested further research using test mean scores for each teacher experience group rather than pass rates. They cited the unique opportunity for research with the large Texas data set.

Clotfelter, Ladd, and Vigdor's 2004 study revisited the question of teacher effectiveness with a focus on whether observable teacher characteristics such as years of experience and scores on licensure tests were associated with higher gains in student achievement. They based their research on the education production functions research of the past half century.

The researchers devised a three-component strategy for estimating teacher effects in the presence of other variables such as teacher sorting and shopping (the tendency for more highly qualified teachers to migrate to more affluent districts). The strategy involved controlling for student demographics

and characteristics, adding school fixed effects, and restricting the set of schools to those that distribute students randomly across classrooms.

Data for this study were obtained from the North Carolina Education Research Data Center (NCERDC). This data set was chosen because students could be linked to teachers and it included test results from 117 diverse school districts. Clotfelter et al. (2004) confined their study to fifth-grade students and their teachers. Personnel records providing the teacher data were retrieved from the state-maintained archive of personnel records. The study report did not indicate the exact number of student records examined.

Following the education production functions tradition, the researchers focused on observable, measurable teacher qualification characteristics, such as years of experience and scores on licensure exams as independent variables. Student achievement, as measured by test scores, was used at the dependent variable. Fifth grade math and reading test scores were standardized in each regression to have mean zero and standard deviation of one.

Clotfelter et al. (2004) found a significant correlation between teacher experience and licensure test scores. Their study determined that the relationship between student achievement and teacher experience was nonlinear, with the peak occurring between 13 and 26 years of experience and the novice teachers associated with the lowest test scores.

After controlling for a rich set of covariates, the authors determined that experience and licensure test scores consistently predicted student achievement. Students benefited approximately one-tenth of a standard deviation on reading and math scores. As a result of their findings, the authors suggested that teacher experience levels should remain a part of pay scale determination.

Summary

Contrary to Hanushek's (1986, 1993) initial findings, researchers between 1996 and 2004 have found that observable characteristics of expert teachers positively correlate to student achievement. Teaching experience (Clotfelter et al., 2004; Goldhaber & Brewer, 1996; Greenwald et al., 1996; McElroy & Pai, 2003; Rivkin et al. 2002; Rowan, 2002) and certification (Clotfelter et al., 2004; Rowan, 2002; Wayne, 2003) were both identified as characteristics of teachers with high student achievement or test score gains, Additionally, overall quality of teachers was found to be a determinant of student achievement (Darling-Hammond, 2000; Provasnik, 2003). Findings of the researchers in this sub-section are summarized in Table 4.

Table 4

Relating Teacher Characteristics to Student Achievement

Researcher	Sample Size	Characteristics Affecting Achievement
Greenwald, Hedges & Laine (1996)	Meta-analysis 60 studies	Teacher education level, teacher ability, teacher experience, per pupil expenditures
Darling-Hammond (2000)	5,600 school districts	Teacher certification
Goldhaber & Brewer (1996)	5,149 students	Teacher preparation, teacher questioning, curriculum control, teacher experience, teacher gender, subject matter knowledge
Okpala, Smith, Jones & Ellis (2000)	4,256 students	Class size, teacher education level, teacher experience, poverty
Rivkin, Hanushek, & Kain (2002)	200,000 students	Teacher experience, teacher quality, school characteristics

Table 4 (continued).

Relating Teacher Characteristics to Student Achievement

Researcher	Sample Size	Characteristics Affecting Achievement
Rowan, Correnti, & Miller (2002)	Thousands of students	Teacher experience
Provasnik & Stearns (2003	12,000 students	Teacher quality
Wayne & Youngs (2003)	Meta-analysis 21 studies	Teacher certification
McElroy & Pai (2003)	16,718 students	Teacher experience
Clotfelter, Ladd, & Vigdor (2004)	117 school districts	Teacher experience, teacher licensure

Teacher Behaviors (Process Variables)

The decade of the 1960s was marked with Coleman, Campbell, Hobson, McPartland, Mood, & Weinfeld's (1966) publication of the Coleman Report titled *Equality of Educational Opportunity* which concluded that the quality of a student's schooling accounts for only about ten percent of the variance in

the student's achievement. Jencks et al. (1972) corroborated the results of Coleman et al. in 1972 with *Inequity: A Reassessment of the Effects of Family and Schools in America.* Together these two publications painted a dismal picture of the education system and its influence on student achievement and while discouraging some from further research, their findings challenged others to prove them wrong.

With the advent of the 70s, researchers began to look at classroom teachers and how their actions contributed to student learning. Brophy and Good (1986) reviewed hundreds of these studies and concluded: "The myth that teachers do not make a difference in student learning has been refuted" (p. 370).

Several of the teacher effect studies were reviewed in the preceding subsection on *teacher characteristics* or *presage variables.* However, researchers have had difficulty agreeing on specific *teacher behaviors* that make up the *process variables.* The investigations included the instructional interactions between students and teachers, instructional strategies, teacher professional development, and even the affective dispositions of teachers in an effort to make connections between student achievement and teacher expertise. In this sub-section I review literature on process variables and their relationship to student achievement.

Through the 70s and 80s most of the educational research on teacher effect was specific in focus. Each study investigated one particular strategy or

aspect of the teacher-learner relationship. Alvermann and Boothby (1986) documented the effectiveness of the use of graphic organizers. Anderson and Armbruster (1986) focused on the value of note-taking. Deci (1971) studied the effects of rewards on intrinsic motivation. Hewson and Hewson (1983) testified to the importance of connecting to prior knowledge. Kounin (1970) documented the importance of classroom management, coining the term "withitness" to describe teacher awareness and interactions with the classroom. Kintsch (1979) focused on the importance of teacher modeling. Many others investigated other aspects of the teaching process and how it correlated with student achievement.

Each of these contributed to the body of knowledge, but not until Brophy and Good (1986) synthesized this growing body did the larger focus on teacher behaviors or process variables materialize. Brophy and Good's study of two decades of research (250 references) concluded that there is a correlation between teacher behavior and student achievement. Findings included:

1. Engagement rates depend on teachers' ability to organize and manage the classroom
2. Achievement is maximized with overviews, advance organizers, and review
3. Teachers' questions should not have a right or wrong answer
4. Pausing after questioning creates think time

5. Involve all students
6. Address student questions
7. Appropriate instruction varies with the teacher's objectives

Porter and Brophy (1988) conducted another synthesis of research findings on the effects of classroom processes on student achievement. In this study they developed a model of good teaching based on ten years of research from the studies of the Institute for Research on Teaching. The model includes teachers' personal experiences, routines, preactive planning, teachers' knowledge and convictions regarding content, pedagogy and student needs, teacher reflections, interactive decisions, and student responses. Their model illustrated that "teaching is highly complex, containing many points for possible breakdown or error. The best teachers negotiate their way through this complexity by attending to each relevant factor" (p. 75).

Like Brophy and Good (1986) and Porter and Brophy (1988), Wang, Haertel, and Walberg (1990) synthesized the research findings of others. They reviewed 179 handbook chapters, 91 research syntheses, and 61 educational researcher surveys to compile 11,000 statistical findings. Wang et al. conducted their meta-analysis using a 28-category conceptual framework to summarize their results. When the 28 categories were ranked by their average rate of influence, classroom management and student/teacher social interactions were in the top five influencers. They found that the amount of time a teacher spends on a topic and the quality of

the social interactions with students have a greater impact on learning than policies adopted by the school, district, or state. Overall, their findings supported "renewed emphasis on psychological, instructional, and contextual differences . . . Unless reorganization and restructuring strongly affect the direct determinants of learning, they offer little hope of substantial improvement" (p. 79).

Cohen and Hill (2000) investigated such a reorganization and restructuring effort in California in which teacher-learning opportunities in mathematics professional development were related to classroom practices in mathematics instruction recommended by the California mathematics framework. The researchers used data from a 1994 survey of California elementary school teachers and 1994 student California Learning Assessment System (CLAS) scores. Using these data sources they examined the influence of assessment, curriculum, and professional development on teacher practice and student achievement.

Cohen and Hill (2000) designed a survey instrument which they administered to four (grade 2-5) teachers from 250 schools (N= 1,000). Because the number of students in each system varied greatly, the researchers stratified the districts by student population and drew the samples proportionate to district size. Because some schools did not support enough teachers for the samples, the final sample of participants was 975 (N = 975). The survey asked teachers to self-report on their classroom practice

using the dimensions advocated by the California Frameworks and how their teaching compared with conventional practice.

The independent variables included attendance at curriculum-centered workshop and time in the workshop. Dependent variables were framework-related practice and student performance on fourth grade CLAS.

The authors used the Ordinary Least Squares (OLS) regressions and found that the content of teachers' professional development made a difference in practice. Specifically, teachers who attended a weeklong student curriculum workshop scored higher on the framework practice scale and reported fewer conventional practices. Attendance at the workshop not only appeared to increase innovative practice but seemed to decrease conventional practice.

Next, the researchers merged student scores on the 1994 fourth-grade mathematics CLAS onto the school files data set to determine if changes in teacher practice led to improvements in student performance. Findings indicated that schools in which teachers reported classroom practices based on the California Framework had higher fourth-grade math scores. Cohen and Hill (2000) also concluded that time spent in math curriculum workshops correlated positively with fourth-grade math scores. Generally, their findings suggested that when educational improvement was based on learning and teaching academic content, and this training overlapped with curriculum and

assessment, teaching performance and student achievement were likely to improve. They wrote that:

> . . . policy makers and practitioners would be well advised to more solidly ground teachers' professional education in deeper knowledge of the student curriculum, or that it would be wise, when new curricula and assessments are being designed, to make much more adequate provision for teachers to acquaint themselves with and learn from them. (p. 332)

Cohen and Hill's (2000) findings reflected those of Greenwald et al. (1996), Darling-Hammond (2000), and Okpala et al. (2000) in their emphasis on the value of a highly qualified teaching staff. Their findings were more specific as to the definition of quality. While the other three studies addressed teacher ability, education, experience, Cohen and Hill focused on teacher practice by identifying the practice and quantifying the results. They also reinforced the findings of Wang, Haertel, and Walberg (1990) by showing the success of a reorganization initiative that was designed to affect a direct determinant of student learning – instructional strategies.

Instructional strategies were also the focus of Marzano, Pickering, and Pollock (2001) when they completed a meta-analysis of over 1,400 studies spanning the past 30 years. His goal was to identify those instructional strategies that had a high probability of enhancing student achievement for students in all grade levels and subject areas. Marzano et al.'s work resulted in a list of strategies in order of their effect on student learning ranging from a 45 percentile point gain to a 22 percentile point gain:

1. Identifying similarities and differences
2. Summarizing and note taking
3. Reinforcing effort and providing recognition
4. Homework and practice
5. Nonlinguistic representations
6. Cooperative learning
7. Setting objectives and providing feedback
8. Generalizing and testing hypotheses
9. Questions, cues, and advance organizers (Marzano, Pickering, & Pollock, 2001, p. 7)

Wenglinsky (2002) also explored the link between classroom practices or strategies and student achievement. Additionally, he studied two other aspects of teaching: teacher professional development and teacher background characteristics. The study used data on 7,146 eighth-grade students who participated in the 1996 National Assessment of Educational Progress (NAEP) in mathematics.

The researchers used the NAEP academic measure for several reasons. First, in addition to student achievement, it measured three aspects of teacher quality: (a) teacher classroom practices, (b) teacher professional development, and (c) educational attainment. Second, it measured the teacher's educational attainment level, whether the teacher majored or minored in the relevant subject area, and the teacher's years of experience.

Wenglinsky (2002) used multilevel structural equation modeling (MSEM) because this technique had the capability to distinguish between students and schools, took measurement error into account, and estimated interrelationships among independent variables. Wenglinsky took socio-

economic status and class size into account when he analyzed the data. The MSEM produced t-scores of the indirect effects and statistics that measured the overall goodness of fit.

The study identified five variables positively associated with achievement: (a) teacher major, (b) professional development in higher-order thinking skills, (c) professional development in diversity, (d) hand[s]-on learning, and (f) higher-order thinking skills. These findings supported the findings of Cohen and Hill (2000) in the connection among professional development, teacher use of the training, and appropriate assessment.

In sum, Wenglinsky (2002) found that "schools matter because they provide a platform for active, as opposed to passive, teachers" (p. 24). He found that schools with a critical mass of teachers who possess the five identified variables can help students reach higher levels of academic performance.

Stronge (2003) completed another review and synthesis of the research related to effective teaching. His goal was to summarize research results accumulated over several decades to define specific teacher behaviors that contribute to student academic achievement and other measures of effectiveness. Stronge was not specific as to the number of studies included in this synthesis. However, there are approximately 100 citations of research studies included in his work.

Stronge (2003) organized his results into 30 categories and then further summarized these 30 categories into six themes in no particular order of importance. These themes included:

1. Prerequisites of effective teaching (personal traits, not education or certification)
2. The teacher as a person
3. Classroom management and organization
4. Organizing for instruction
5. Implementing instruction
6. Monitoring student progress and potential

Although none of the studies in this subsection exactly match each other in areas studied or findings, there are similar threads that run through the group. Stronge's (2003) synthesis summarizes and pulls the others into a cohesive unit. It ties all the strands of teacher expertise to student academic achievement.

Summary

The findings in this sub-section illustrate the great variety of variables at work in today's classroom. The myriad studies addressing these variables have led to several meta-analyses reviewed in this subsection. Instead of being the final word, The Coleman Study (1966) seemed to spawn a stream of studies documenting the value of teachers to the success of schools and students. Brophy (Brophy & Good, 1986; Porter & Brophy, 1988) played an

important role in defining the behaviors of good teachers through his meta-analysis and model depicting the complexities of the teaching process. Cohen and Hill (2000) and Wenglinsky (2002) contributed to the knowledge base by identifying professional development. Wang et al. (1993), Marzano et al. (2001) and Stronge (2003) also completed more recent syntheses of research on teacher behaviors. Findings of all the in this sub-section are summarized in Table 5.

Table 5

Relating Teacher Behaviors to Student Achievement

Researcher	Sample Size	Behaviors Affecting Achievement
Brophy & Good (1986)	Meta-analysis 250 studies	Teacher organization, overviews, advance organizers, review, questioning, student involvement, varying instruction

Table 5 (continued).

Relating Teacher Behaviors to Student Achievement

Researcher	Sample Size	Behaviors Affecting Achievement
Porter & Brophy (1988)	Synthesis	Teachers' personal experiences, routines, preactive planning, content knowledge, pedagogical knowledge, reflections, interactive decisions
Wang, Haertel, & Walbert (1990)	Synthesis	Classroom management, student/teacher social interactions
Cohen & Hill (2000)	1,000 students	Professional development content, adherence to curriculum
Marzano, Pickering & Pollock (2001)	Meta-analysis	Similarities/differences, summarizing, praise, practice, nonlinguistic representations, cooperative learning, objectives, generalizing, questioning

Table 5 (continued).

Relating Teacher Behaviors to Student Achievement

Researcher	Sample Size	Behaviors Affecting Achievement
Wenglinsky (2002)	7,146 students	content knowledge, professional development in higher-order thinking skills and diversity, hands-on learning, use of higher-order thinking skills
Stronge (2003)	Synthesis	Personal traits, classroom management, organization, implementing instruction, monitoring student progress

Presage and Process Variables

Ferguson (1991) did not limit his study to presage or process variables, focusing instead on the extreme importance of good teachers. This focus included the influence of both. He studied 900 of 1,063 Texas school districts and measured teacher expertise by scores on a licensing examination, master's degree, and experience. The study used information gathered from

student and teacher tests in Texas in the late 1980s including more than 2,400,000 students. The sample is over four times as large as that used in the largest previous study, the "Coleman Report" of almost 30 years ago. Ferguson's findings include.

1. Good teachers are the most important factors in good education.
2. After taking all those social and economic factors into account, the teachers' test scores and years of experience are the most important factors in student test scores. Higher salaries draw smarter people into the teaching profession and keep them there longer.
3. In primary schools, teachers with five or more years of experience get the best results. In high schools, teachers take nine years to reach their best performance.
4. For every 10% increase in the number of experienced high school teachers, the dropout rate goes down four percent and the number of students taking the Scholastic Aptitude Test to qualify for college goes up by three.

Sanders and Rivers (1996) also did not limit their study to either process or presage variables individually. Like Ferguson (1991) their study focused on the overall quality of teachers and their effect on student achievement. They examined the estimates of cumulative teacher effects in mathematics from grades three through five using the data from two of Tennessee's larger metropolitan systems. Data for this investigation included

a cohort of students who were second graders in 1991-92, third graders in 1992-93, and fourth graders in 1994-95. Exact sample size was not included.

Sanders and Rivers (1996) used the sequence of teachers as the independent variable effecting student achievement. Scale score gains on a norm-referenced test comprised the dependent variable for this study.

The researchers used a multivariate longitudinal analysis of scale score gains to produce estimates of teacher effects. After the teacher effects were obtained for each grade level, teachers in each grade level were grouped into quintiles. Teachers exhibiting the highest degree of effectiveness were placed in the fifth quintile and teachers with the lowest degree of effectiveness were in the first quintile. By encoding individual student records with teacher quintiles and using crosstabs, the researchers traced the progress of these students through identified sequences of teacher effectiveness. Thus, they determined whether teachers from previous years affected following year scores.

Sanders and Rivers (1996) credited teacher sequence with 50 percentile point differences in student achievement. They also found that teacher effects on student achievement were additive and cumulative with little evidence of compensatory effects.

The study strongly suggests the presence of cumulative effects of teachers on student achievement. Groups of students with comparable abilities and initial achievement levels may have vastly different academic

outcomes as a result of the sequence of teachers to which they are assigned. These analyses also suggest that the teacher effects are both addictive and cumulative with little evidence of compensatory effects of more effective teachers in later grades. The residual effects of both very effective and ineffective teachers were measurable two years later, regardless of the effectiveness of teachers in later grades.

Wright, Horn, and Sanders (1997) also analyzed 100,000 student achievement scores across hundreds of schools in Tennessee. Their conclusion was

> . . . the most important factor affecting student learning is the teacher. In addition, the results show wide variation in effectiveness among teachers. The immediate and clear implication of this finding is that seemingly more can be done to improve education by improving the effectiveness of teachers than by any other single factor. Effective teachers appear to be effective with students of all achievement levels, regardless of the level of heterogeneity in their classrooms. If the teacher is ineffective, students under the teacher's tutelage will show inadequate progress academically regardless of how similar or different they are regarding their academic achievement. (p. 63)

Summary

While each of these three studies had a slightly different perspective, they combine to give readers a more complete picture of effective teachers as they affect student achievement. Ferguson (1991) and Sanders (Sanders & Rivers, 1996; Wright, Horn, & Sanders 1997) give an overall picture of the value and impact of good teachers. Of special importance are the Sanders studies (Sanders & Rivers, 1996; Wright Horn, & Sanders, 1999) because of

their findings on the residual effect of ineffective teachers. Findings of all the researchers in this sub-section are summarized in Table 6.

Table 6

Relating Presage and Process Variables to Student Achievement

Researcher	Sample Size	Presage and Process Variables Affecting Achievement
Ferguson (1991)	900 school systems	Teacher quality, teacher test scores, teacher experience
Sanders & Rivers (1996)	Large (Memphis schools)	Teacher sequence (effective and ineffective teachers)
Wright, Horn, & (1997)	100,000 students	Teacher quality

Summary and Research Problem

Thirty years of research on teacher expertise and its specific effect on student achievement have yielded a rich field of inquiry. This literature review examined the findings on the differences between novice and expert teachers and the implications these differences have for the nation's students. It highlighted the cognitive processes, subject matter expertise, and variety of

types of knowledge that expert teachers possess (Leinhardt & Greeno, 1986; Schempp et al., 1998).

These findings were contrasted with the findings on novice teachers' beliefs, skills, and assimilation into their new environment. They were often not prepared to assume the responsibility of student learning (Bullough, 1987; Ralph, 1999). Their need for survival overrode the beliefs systems they had established. Researchers suggested that teacher education programs and school systems establish programs of teacher preparation and induction scaffolding the movement from pre-service teacher to effective teacher.

The literature, comparing novice to expert teachers, reflected a general agreement on the part of the researchers that there was a significant difference in the areas of instructional planning, mental deliberations, strategies, reflections, and responsibility for student learning (Borko & Livingston, 1989; Manning & Payne, 1996; Tochon & Munby, 1993; Westerman, 1991).

Researchers found that observable characteristics of expert teachers positively correlate to student achievement. Teaching experience (Clotfelter et al., 2004; Goldhaber & Brewer, 1996; Greenwald et al., 1996; McElroy & Pai, 2003; Rivkin et al. 2002; Rowan, 2002) and certification (Clotfelter et al., 2004 Rowan, 2002; Wayne, 2003) were both identified as characteristics of teachers with high student achievement or test score gains. Additionally,

overall quality of teachers was found to be a determinant of student achievement (Darling-Hammond, 2000; Provasnik, 2003).

Further research on teacher behaviors provided insights into the effects of expert teachers on student achievement. Teaching strategies, teacher dispositions, and professional development are positively correlated to student achievement. Several meta-analyses in the past decade have brought cohesiveness to the vast number of studies on teacher behaviors (Brophy & Good, 1986; Marzano et al., 2001; Porter & Brophy, 1988; Stronge, 2003; Wang et al., 1993)

The one recurring theme in almost every study was *experience.* No other variable was tested more often. No other selection criterion was used more often. No other finding was listed more often than teacher experience. All of the teacher quality studies are not in complete agreement on the linear relationship between teacher experience and expertise. However, all the studies on novice teachers and their comparison to expert teachers agree that expertise is not a term synonymous with novice. The frequent use of teaching experience as a variable and the frequency of its identification as an influencing finding point to its importance in future research.

The research summarized in this chapter highlights the need for quality before accountability. Local school systems must have a quality teaching force in place before they can meet the accountability mandates set forth by state legislatures. Novice teachers are assigned the same numbers of

instructional objectives to teach the same numbers of students, in the same time frame as the more experienced teachers. The inconsistency of these expectations is in direct contrast to their level of preparation and expertise.

Just as the majority of teacher quality research has its roots in school resources research, so does accountability grow from cost accounting. Comparing the output efficiency of a school system, school, or teacher is a complex idea with myriad possibilities, with a difficult-to-define variable of teacher expertise near the center.

Of special importance to this study is the finding by Sanders and Rivers (1996) on the cumulative and additive effects of an ineffective teacher. If the findings of the researchers in this review were accurate, novice teachers lack many of the qualities of effective teachers. What are the implications of these findings to schools and systems that employ large numbers of novice teachers? Does the effect of a teacher's lack of expertise follow students throughout their school career?

With the current focus on teacher accountability, what is the role of the school system in supporting the novice teacher? What programs can local school systems implement to offset this achievement deficit for students in novice teachers' classrooms?

Research Purpose

The purpose of this descriptive study was to investigate the relationship between years of teaching experience and future student

achievement in reading. Reading was chosen as the subject area for the study because most academic areas display some dependence on success in this skill. Coley and Coleman (2004) posited that effective reading and literacy instruction are keys to educational success and form a critical component in efforts to close the gaps in student achievement between social classes and between racial/ethnic groups. Snow, Burns and Griffin (1998), in their study for the National Research Council, concluded that "quality classroom instruction in kindergarten and the primary grades is the single best weapon against reading failure" (p. 343). Cognizance of the factors affecting this process would be invaluable to school leaders.

CHAPTER III

METHODOLOGY

The review of the literature began with the 1980's research on characteristics of effective teachers and progressed into the 21st century research on the achievement effects of these teachers. As the journey progressed through time, it also progressed through the teacher career cycle from novice to expert. After 25 years of research on effective teachers, novice teachers, the differences between the two groups, and the student achievement of effective teachers, questions remain. Berliner (1986) and other researchers, through the last three decades, have identified developmental stages of teachers, but few have attempted to link these stages to student achievement.

This study examined the annual achievement test NCE scores of students in a suburban/rural system in middle Tennessee to answer these questions presented in Chapter I.

1. Is there a statistically significant difference between the reading achievement, measured two years later, of students taught by a novice teacher in primary grades (kindergarten, one, and two) and the state mean for the appropriate grade level?

2. Is there a statistically significant difference between the reading achievement, measured two years later, of students taught reading by a second or third year teacher in primary grades (kindergarten, one, and two) and the state mean for the appropriate grade level?
3. Is there a statistically significant difference between the reading achievement, measured two years later, of students taught reading by a teacher with three or more years of teaching experience (teachers in their fourth or greater year of teaching) in primary grades (kindergarten, one, and two) and the state mean for the appropriate grade level?
4. Is there a statistically significant difference among the two-year-later reading achievement of the three groups of students in questions one, two, and three?
5. The study was designed to assess teacher growth from their first year through their second and third and to assess the degree to which the future achievement of students is related to their time under the tutelage of a novice teacher two years prior. The study was narrow in its objective, focusing on a small number of students and their progress over a two-year period relating to the experience level of a primary reading teacher. It was grounded in the research on teacher development. How do teachers develop and what is the result of their development?

Practicing teachers learn about teaching in many ways. They learn from their own experiences with what works and what does not work in specific contexts. This pedagogical reasoning (Wilson, Shulman, & Richert, 1987) allows teachers to gain knowledge and understanding of their students, schools, curriculum, and instructional methods by living daily practical experiments. A second learning method for practicing teachers is in their interactions with more experienced teachers during formal and informal mentorship. Third, teachers learn through professional development provided in their work environment. A fourth teacher-learning possibility is through advanced degree programs the teacher might be pursuing. Finally, teachers learn about teaching in arenas outside their professional lives. Personal growth, maturity, and intellectual development contribute to their performance as a teacher (Bransford, 1999). These learning opportunities take place over a number of years and contribute to the expertise of effective teachers.

A 2002 study conducted in this school system compared student gains for teacher experience groups and revealed lower achievement test gains for students of beginning teachers than other students. The study was based on student scale score gains at fourth grade and higher. The national norm gain on the TerraNova Achievement Test from the end of third grade to the end of fourth grade reading was 12 scale scores. Table 7 shows the mean scale score

gain for students in the fourth grade disaggregated by the experience level of their teacher.

Table 7

Fourth Grade Mean Reading Scale Score Gain (2002)

Teaching Experience	Mean Scale Score Gain	National Scale Score Norm Gain	% of National Scale Score Norm Gain
0-2 Years	9.9	12	83%
3-14 Years	13.8	12	115%
15-32 Years	14.3	12	119%

In 2002, the Tennessee grading system for schools and systems required a gain of 12 scale scores (the national norm) in fourth grade reading for a letter grade of "C." The 9.9 scale score growth (83 % of 12) exhibited by the novice teacher group would have received a letter grade of "F" from the state. The other two experience level groups would have earned an "A." This discrepancy has the potential to affect both school and system grades and the success of students in these classrooms. The state grading scale for fourth grade reading at that date (2002) is shown in Table 8.

This disparity in gains among the groups is greatly exacerbated by the large number of inexperienced teachers. In the school system in this study, over 33% of the staff was in the zero to two-year experience level group.

Table 8

Tennessee Grading Scale for Fourth Grade Student Gains (2002)

Grade	Status	Reading
A	Exemplary	> 115% of 12 Scale Score gains
B	Above Average	106% - 115% of 12 Scale Score gains
C	Average	95% - 105% of 12 Scale Score gains
D	Below Average	85% - 94% of 12 Scale Score gains
F	Deficient	< 85% of 12 Scale Score gains

In Tennessee, schools and systems also receive grades on student achievement, in addition to the gains scores. NCE achievement scores were examined for the three experience categories included in this study. The results are displayed in Table 9. Although the scores do not seem to differ greatly, a one-point difference can affect letter grades (A, B, C, D, and F) assigned to each school by the state department of education.

Table 9

Normal Curve Equivalent (NCE) Scores for Experience Level Groups in Grades Two, Three, and Four

Experience Level	2002-2004 Group 3-Year Mean NCE	2002-2004 System 3-Year Mean NCE	2001-2003 State 3-Year Mean NCE
Novice	56.0	56.4	52
Years 1 & 2	57.2	Includes all experience levels, Special Education, and Title I	Includes all experience levels, Special Education, and Title I
Years 3 or more	58.8		

Note: Beginning in 2004, Tennessee no longer provides a state mean. However, the state mean for the previous three years was 52 NCEs, and the three-year mean for 2003 was 52 NCEs.

The grading scale for achievement is found in Table 10. This scale is peculiar to the state of Tennessee and does not meet accepted standards for average, above average, and below average.

Table 10

Tennessee Grading Scale for Student Achievement

Grade	Status	Reading
A	Exemplary	57-99 NCE Score
B	Above Average	52-56 NCE Score
C	Average	46-51 NCE Score
D	Below Average	41-45 NCE Score
F	Deficient	0-40 NCE Score

Does this lower achievement and gain rate of students taught by novice teachers persist in future years? This exploratory study followed students taught reading by each of the three experience level groups to determine if there was a persistent and residual effect on achievement resulting from primary reading instruction. The study was planned to collect reading test data two years after the primary reading instruction so that students would have two full years to practice learned skills and acquire new ones. The two-year period allowed the researcher a clear picture of the long-term outcome. The two-year interval also allowed students to experience two

more teachers, to ensure that they receive every opportunity to overcome any long-term effect of the novice teacher.

Chapter III is divided into six sections containing information pertinent to the methodological tools used in the culminating analysis for this study. The first section of this chapter contains a detailed description of the *subjects* of the study, the number of participants, how the number was derived, the sampling method, a detailed description of the community involved in the study, a detailed demographic description of the sample, and the rationale for the subject selection.

The second section in Chapter III describes the *design* of the study; which data were collected, and how each piece of data relates to the study. This section also contains a description of the dependent and independent variables and the measurement of each variable. The third section in Chapter III includes a discussion of the test *instrument* administered to measure student achievement including the format and the reliability of the instrument. This section also includes an explanation of the scores generated by the test, and used in the study.

Section four describes the data collection *procedures* and details the protection of the rights of participants during the study. Section five describes the *data analysis*. This section identifies the statistical techniques used to test each of the research questions in the study. The final section identifies *assumptions and limitations* present in the study.

Subjects

This study used the purposeful sampling process to select participants. Purposeful sampling is the "selection of individuals/groups based on specific questions/purposes of the research in lieu of random sampling and on the basis of information available about these individuals/groups" (Tashakkori & Teddlie, 1998, p. 76). In this instance, the purpose was guided by teacher experience. The students (N= 4,588) were selected from six elementary schools in a relatively small rural/suburban school system with a total student population of 7,000 students in grades kindergarten through 12.

The disproportionately large number of primary students results from the fact that some teachers fit different teacher categories in different years of the study. For example, Teacher A's students may have been studied during the teacher's novice year, again the second year when they were in the classroom of a second-year teacher, and again the third year when they were in the classroom of a third-year teacher. The student population studied included all kindergarten through second-grade students taught reading or reading readiness by a teacher who met these qualifications:

1. taught kindergarten, first grade, or second grade reading in a regular classroom the school system in either school year 1999-2000, 2000-2001, or 2001-2002
2. was licensed and endorsed to teach reading in grades kindergarten through two

3. taught for the full school year excluding the sick leave, personal, or professional days provided by Tennessee law

All kindergarten, first-, and second-grade students for school years 1999-2000, 2000-2001, and 2001-2002 were selected for the initial part of the study. Because this study was longitudinal, not all students remained in the study. The student scores used in the study were those of students who were still in the system and participated in the annual achievement test two years after the date of initial participant data collection. For example, kindergartners in 1999-2000 were the subject of achievement reading test score analysis at the end of the second grade (2001-2002) and first graders in 2000-2001 were the subject of achievement reading test score analysis at the end of the third grade (2002-2003). Reading was used as the subject area for the study because most academic areas display some dependence on success in this skill. The United States Department of Education described the importance of reading in this manner.

> Countless new doors are opened when children become good readers early in life. . . Young, capable readers can take greater advantage of school opportunities, and develop invaluable confidence in their own abilities. Plus, reading success leads directly to success in other subjects such as social studies, math, and science. In the long term, students who cannot read well are much more likely to drop out of school and be limited to lower-paying jobs throughout their lifetimes. Reading is undeniably the foundation for success in society. Reading must come first. (U. S. Department of Education, n. d.)

Because reading is vitally important to success in all content areas, principals need all the information possible as they make placement decisions

within their schools. Reading instruction should be the responsibility of the most expert reading teachers available.

Focusing the study on one system allowed the researcher to compare the scores of these three groups influenced by the same system history, administration, leadership, and decision-making process (Meyers, Meyers, & Gelzheiser, 2001). The school system used in the study was selected because of its close proximity to the researcher and its geographical location. Surrounded by three larger urban/suburban systems with higher pay scales, the system has a history of a high teacher turnover rate. This creates an environment in which novice teachers serve a high percentage of system students. Approximately 33% of the elementary teachers in any school year have fewer than three years of teaching experience.

Table 11 lists the classroom variables which often affect achievement and compares the six schools in each area.

Table 11

Classroom Variables

	School 1	School 2	School 3	School 4	School 5	School 6
Teacher Education Level						
MA	24%	23%	32%	11%	20%	5%
BS	76%	77%	68%	89%	80%	95%
Texts in Use	All schools use the texts adopted by the system					
Ethnic Diversity						
Caucasian	95.6%	98%	98.6%	97.7%	97.9%	95.1%
Other	4.4%	2%	1.4%	2.3%	2.2%	4.9%
Funding	All schools funded on a per-pupil basis by system					
Poverty Level (Free/Reduced)	52.9%	45.7%	19%	22.8%	21.1%	34.2%
Mean Experience						
Level (years)	14.6	11.6	13.7	10.6	12.9	10.6
Title I Service	Yes	Yes	No	No	No	Yes
3-Year Mean						
Achievement Score	52	52	64	48	64	52
Report Card Grade	B	B	A	C	A	B

Source: School system student management system and annual report card

All six schools were comparable in some of the variables examined and disparate in others. The poverty rates varied greatly and were negatively correlated to achievement in schools 1, 2, 3, and 5. Scores in schools 4 and 6, however, were not as easily explained. Not discernible in Table 11 is the fact that there has been an extremely large teacher turnover in School 4 in the past two years, including a new principal.

Use of a locally enhanced state curriculum and texts adopted for system-wide use provided some consistency in instruction. The poverty rates at the schools varied from approximately 19% to 52.9%, using *free or reduced* lunch percentages to determine poverty. The percent of students who qualify for free or reduced priced lunch is used in Tennessee to qualify schools for Title I remedial reading and math services. The three schools with the higher poverty rates received Title I services.

Schools 1 and 2 (the highest poverty schools) had two of the three highest percentages of teachers with advanced degrees. School 1 (highest poverty) had the highest average in teacher years of experience. In most cases, the schools with the highest averages in teacher years of experience had the highest achievement scores.

The six schools included in the study ranged from 48 to 64 NCEs on their achievement test scores. While the scores were not closely comparable, in broader terms, outside the Tennessee framework, they were all within the accepted average range. According to the publisher of the achievement test

used, scores below 34 are considered below average; scores between 34 and 65 are considered in the average range; and scores above 65 are in the above average range (CTB/McGraw, n.d.).

The school system studied is a countywide system, encompassing all students in the county with the exception of one small private school. According to the 2000 census, the per capita income for the county was $18,882, which is slightly below the state average of $19,393. However, the median household income was $45,836 compared to the Tennessee average of $36,360 at that time (United States Census, 2000). Employment figures for the third quarter of 2004 indicated that the county's unemployment rate was 3.9%, which was much lower than the state average of 4.8% (Federal Deposit, 2004).

Pupil-teacher ratios were lower than the state requirements. All elementary teachers were licensed and endorsed in their subject area or grade level by the state of Tennessee. There were no provisional or emergency licensed teachers. The only exceptions to this are approximately five teachers each year who transfer into the system from another state where the PRAXIS test is not required. The PRAXIS Series from Educational Testing Service (ETS) are tests designed to be used principally in connection with other criteria by state authorities for the purpose of licensing education professionals. PRAXIS tests are used for credentialing purposes and focus on a candidate's current skill, knowledge, or competency in a particular domain.

Tennessee requires that teaching candidates successfully complete a PRAXIS test in their specific domain before becoming licensed and endorsed. The teachers who enter the system with experience elsewhere and lacking the PRAXIS are given one year for successful completion of the domain test. This group of teachers was not included in the study. Additionally, teachers who serve exceptional populations solely (i.e., Title I), were not included.

All system schools were accredited by the Southern Association of Colleges and Schools and had been for nine years or more. The system surpassed state personnel requirements providing teaching assistants in the primary grades, school counselors, physical education teachers, and music teachers in all elementary schools. All the schools in this study made adequate yearly progress in 2003-2004 as required by *No Child Left Behind* (P.L. 107-110) (NCLB).

However, several schools in the study received Ds and Fs in reading on their most recent value-added report for 2003-2004, released to systems on January 3, 2005. Tennessee's value-added report is based on student achievement growth from one year to the next. Until school year, 2003-2004 value-added scores were calculated by measuring scale score gains. In 2004, the Tennessee Department of Education revised the process to reflect school performance on the Criterion Referenced Test (CRT) portion of TerraNova as evidenced by improvement in NCE scores.

The minority population, including all racial and ethnic groups, totaled approximately 1.5% of the total population. Because of this low percentage of minority students, the data were not disaggregated by race or ethnicity. The most recent Consolidated Plan for the system showed no significant difference in the performance of males and females. Therefore, the data were not disaggregated by gender. Table 12 gives approximate numbers of student NCE scores used in the study. It illustrates that there were more than 2,800 NCE scores collected. This is far greater than the actual number of students. However, data were collected for all students at least twice.

Table 12

Description and Estimated Numbers of Students in Two-year-Later Comparison

Grade level	Teacher experience		
	Novice	1-2 years	3 Years or more
K	100	180	560
1	120	200	640
2	140	240	680

Design

This exploratory study employed the examination of archived longitudinal data collected over a five-year period beginning in the spring of

school year 1999-2000 and ending in the spring of school year 2003-2004. Student attendance data were examined for years 1999-00, 2000-01, and 2001-02 to identify kindergarten through second-grade students taught by teachers in each of the three experience categories. Two-year later test data for these identified students were collected and analyzed. The study contained nine comparisons of grade level and teacher categories. Tables 13, 14, and 15 illustrate the data collection comparisons for questions one, two, and three.

Table 13

One Sample *t*-test Comparisons to Address Research Question 1

Novice Teacher K	Years of K	District Test Data for for grade 2	State Data for grade 2	One Sample *t* comparisons
	2000	2002	Mean 2002 = 52	
	2001	2003	Mean 2003 = 52	*t* 1
	2002	2004	Mean 2004 = 52	
Novice Teacher Grade 1	Years of Grade 1	District Test Data for for grade 3	State Data for grade 3	One Sample *t* comparisons
	2000	2002	Mean 2002 = 52	
	2001	2003	Mean 2003 = 52	*t* 2
	2002	2004	Mean 2004 = 52	
Novice Teacher Grade 2	Years of Grade 2	District Test Data for for grade 4	State Data for grade 4	One Sample *t* comparisons
	2000	2002	Mean 2002 = 52	
	2001	2003	Mean 2003 = 52	*t* 3
	2002	2004	Mean 2004 = 52	

Table 14

One Sample *t*-test Comparisons to Address Research Question 2

2nd or 3rd Year Teacher for K	Years of K	District Test Data for Grade 2	State Data for grade 2	One Sample *t* comparisons
	2000	2002	Mean 2002 = 52	
	2001	2003	Mean 2003 = 52	*t* 4
	2002	2004	Mean 2004 = 52	
2nd or 3rd Year Teacher for Grade 1	Year of Grade 1	District Test Data for Grade 3	State Data for grade 3	One Sample *t* comparisons
	2000	2002	Mean 2002 = 52	
	2001	2003	Mean 2003 = 52	*t* 5
	2002	2004	Mean 2004 = 52	
2nd or 3rd Year Teacher for Grade 2	Year of Grade 2	District Test Data for Grade 4	State Data for grade 4	One Sample *t* comparisons
	2000	2002	Mean 2002 = 52	
	2001	2003	Mean 2003 = 52	*t* 6
	2002	2004	Mean 2004 = 52	

Table 15

One Sample *t*-test Comparisons to Address Research Question 3

3 or more year teacher K	Year of K	District Test Data for for grade 2	State Data for grade 2	One Sample *t* comparisons
	2000	2002	Mean 2002	*t* 7
	2001	2003	Mean 2003	
	2002	2004	Mean 2004	
3 or more year Teacher Grade 1	Years of Grade 1	District Test Data for for grade 3	State Data for grade 3	One Sample *t* comparisons
	2000	2002	Mean 2002	*t* 8
	2001	2003	Mean 2003	
	2002	2004	Mean 2004	
3 or more year Teacher Grade 2	Years of Grade 2	District Test Data for for grade 4	State Data for grade 4	One Sample *t* comparisons
	2000	2002	Mean 2002	*t* 9
	2001	2003	Mean 2003	
	2002	2004	Mean 2004	

Tables 16, 17, and 18 illustrate the timeline for the data collection activities. Data collection took place over a five-year period.

Table 16

Kindergarten Data Collection Activities

1999-00	2000-01	2001-02	2002-03	2003-04
Identify K students of teachers in each of the three categories		Collect 2nd grade test data on these students		
	Identify K students of teachers in each of the three categories		Collect 2nd grade test data on these students	
		Identify K students of teachers in each of the three categories		Collect 2nd grade test data on these students

Table 17

First Grade Data Collection Activities

1999-00	2000-01	2001-02	2002-03	2003-04
Identify 1st grade students of teachers in each of the three categories		Collect 3rd grade test data on these students		
	Identify 1st grade students of teachers in each of the three categories		Collect 3rd grade test data on these students	
		Identify 1st grade students of teachers in each of the three categories		Collect 3rd grade test data on these students

Table 18

Second Grade Data Collection Activities

1999-00	2000-01	2001-02	2002-03	2003-04
Identify 2nd grade students of teachers in each of the three categories		Collect 4th grade test data on these students		
	Identify 2nd grade students of teachers in each of the three categories		Collect 4th grade test data on these students	
		Identify 2nd grade students of teachers in each of the three categories		Collect 4th grade test data on these students

The independent variable was the experience level of the teacher, ranging from novice to thirty-plus years. Kindergarten, first, and second-grade students were disaggregated by years of teaching experience of their

reading teacher in accordance with the criteria outlined in the research questions (less than one year of experience, one to three years of experience, more than three years of experience). This disaggregation was based on efforts to describe the "stages" teachers go through in mastering the art of teaching. Researchers generally posit an initial stage of survival and discovery, a second stage of experimentation and consolidation, and a third stage of mastery and stabilization (Berliner, 1986; Field, 1979; Watts, 1980). These stages are loosely linked to years of experience with stabilization occurring around the time of tenure. In Tennessee, a teacher receives tenure when hired for the fourth year with three years of acceptable evaluations and the recommendation of the director of schools.

The dependent variable was the reading NCE score on the annual achievement test. NCE scores were used because they are on an equal interval scale, allowing them to be manipulated and averaged for comparison purposes. NCE is a score developed for use by the United States Office of Education for use in interpreting the scores of large groups of students. These scores were obtained from the reports submitted to the school district by the state of Tennessee after each spring administration of the TerraNova. Teacher achievement test lists with NCE scores were used as the raw data for all data analysis.

Instrumentation

The TerraNova Complete Battery is the test administered annually to all Tennessee students in grades three through eight and available for kindergarten through second grade at system expense. TerraNova and its predecessor CTBS/4, from the same publisher, have been in use in the state for 14 years. Because the state conducted an equalizing process between the two tests, there are 14 years of continuous compatible data. The system involved in this study tests grades two through eight annually. The test is administered under strict security measures and is scored by the test publishers, CTB/McGraw Hill.

The reading subtest of TerraNova, which was used for data analysis, has a Kuder-Richardson-20 reliability rating range of from .92 to .95 (Missouri, n.d.). A reliability rating measures the extent to which items on a test are homogeneous. Most large-scale tests report reliability coefficients that exceed .80 and often exceed .90 (Rudner & Schafer, 2001).

National percentile rankings are reported by teacher, by student, and by school. Results also include objective mastery, NCE scores, and scale scores. At the time the data for this study were collected, the TerraNova was a norm-referenced test only, designed to compare students tested to national norms. Since the passage of NCLB, the Tennessee version of the test has been revised to the criterion-referenced format. This change had no effect on the reported data from years 1999-2000 through 2002-2003. It eliminated the

state NCE report. Therefore, there was no state 2004 mean NCE score. However, the state NCE for the previous three years was 52 NCEs; making the three-year mean 52. This established a trend that was used for comparison purposes.

Procedures

The researcher applied to the University of Louisville and Western Kentucky University's Institutional Review Boards and the system director of schools for project approval. All three entities granted approval (see appendixes A, B, and C). At no time during this study, or after its completion, was any personally identifiable information for student, teacher, or school made public. Student, teacher, and school confidentiality was protected at all times.

Data collection began with the compilation of a master list of all teachers meeting the qualifications for the study using system personnel records. For questions one, two, and three the master list was disaggregated by years of teaching experience in accordance with the criteria outlined in the research questions. These were (a) less than one year experience (teachers in their first year), (b) one to two years' experience (teachers in their second and third year), and (c) more than two years' experience (teachers in their fourth or higher year). The products of this disaggregation were three lists of teachers. One list included only novice teachers. The second was comprised of

second- and third-year teachers. The third list included teachers with more than three years of experience.

A class list of students for each novice teacher was obtained from Horizons, the system's student management program. NCE scores in reading from the TerraNova administered two years later for each student were obtained from the system's Clarity test score management software from CTB/McGraw Hill. For students who were taught by a novice teacher in kindergarten, second-grade reading NCE scores were recorded. For students who were taught first-grade by a novice teacher, third-grade reading NCE scores were recorded. For students who were taught second grade by a novice teacher, fourth-grade reading NCE scores were recorded.

Once the reading test data for each experience level group for three years was collected, the data were combined. The three years of second-grade test data (2002, 2003, and 2004) were combined for data analysis purposes. The three years of third-grade data (2002, 2003, and 2004) were combined for data analysis purposes. The three years of fourth-grade data (2002, 2003, and 2004) were combined for data analysis purposes.

The same procedure was followed for the other two classifications of teachers: second- and third-year teachers, and teachers with more than three years of experience. The results of this process were three lists of students and their reading NCE scores from two years later. At this point, each teacher name was coded with a number for confidentiality. These lists did not

contain teacher names. The product of this process was a three-part master list showing only a column of NCE scores for each group of students.

Again, all test data came from existing records of educational achievement test scores that were regularly collected as part of normal education activities. All personnel information came from existing records in the system student management system maintained as part of normal education activities.

McCaffrey, Lockwood, Koretz, and Hamilton (2003) evaluated value-added models for teacher accountability. In their discussion of researchers' control for covariates, they addressed the possibility of school or district effects on student achievement and whether such effects are (or should be) omitted from models. In the Tennessee Value Added Assessment Model (TVAAS), Sanders "chose to exclude student covariates rather than possibly underestimate teacher effects" (p. 75).

In contrast, the Dallas accountability system (Webster & Mendro, 1997) made the opposite choice when examining teacher effects and used a complex covariate-adjustment model that included many student variables. While one system chose to err by possibly compounding effects, the other chose to err by possibly overcorrecting. McCaffrey et al. (2003) posit that analysts must decide which potential error is more acceptable.

The authors continued by stating that analyses that attempt to distinguish teacher effects must be able to distinguish those effects from

others at play in the school environment. The direct effect of the school, its programs, and other offerings should not be considered a part of teacher effect. However, removing school effects is not a simple process and it is often difficult to determine what might bias the analysis, much less control for it or remove it.

> Because true teacher effects might be correlated with the characteristics of the students they teach, current [Value Added Models] VAM approaches cannot separate any existing contextual effects from these true teacher effects. Existing research is not sufficient for determining the generalizability of this finding or the severity of the actual problems associated with omitted background variables. . . Furthermore, the extensive simulation studies we conducted imply that some of the findings from the literature would be unlikely to result solely from omitted variables, bias, or confounding, suggesting that these findings are truly the results of teacher effects and not other factors. (McCaffrey et al., p. 113-114)

MaCaffrey et al. (2003) contended there was no current method to "disentangle true teacher effects from student background characteristics in the presence of classroom-level variables and contextual effects and correlation between true teacher effects and student characteristics" (p. 74). With this finding in mind and because this study was conducted in Tennessee, where Sanders studied teacher effects through TVAAS, the researcher elected to not control for student, school, and system variables. Additionally, the schools studied were not greatly disparate in poverty levels and additional programs. The schools serving students with higher poverty levels provided Title I services for these students. This service was the only outwardly discernible difference in the schools.

Data Analysis

Research questions one, two, and three involved the determination of difference between each of the three groups and the state mean. Using the completed lists of data, the reading NCE scores for each experience level group of teachers were analyzed to determine if there was a statistically significant difference between the NCE scores of students taught by the three experience level groups two years earlier and the state mean. Although the students in the study were included in the larger state mean, the numbers were small enough that they did not affect the state mean.

The pooled variance for the system NCE mean scores for each grade were not reported to systems, and are not available for use in data analysis. Therefore, the single sample *t*-test was used to determine if there is statistically significant difference between each experience level group and the state mean. The samples included in the study met the two assumptions required for the single sample *t*-test: a) the values in the sample must consist of independent observations (representative of the population), and b) the population sample must be normal (Gravetter & Wallnau, 2000).

Often, studies with large numbers of *t*-tests on one set of data, analyses results may indicate that statistically significant differences exist due more to a large sample size than to any real differences between student scores. In such studies, we find that setting $\alpha = 0.05$ does not provide sufficient protection against the Type I error. As the number of separate

hypothesis tests increase within a single study, the true α-level for the entire study will be inflated. To adjust for this possibility the researcher adjusted the *t*-test alpha level using the Bonferroni adjustment procedure. This was accomplished by dividing ". . . the alpha level (.05) by the number of *t*-tests to be performed", J. Petrosko (personal communication, June 11, 2005). Because 9 *t*-tests were performed, a Bonferroni-corrected alpha level was used to avoid inflation of Type I error. The alpha level of each test was set at .05/9 = .0055. This completed the process for questions one through three.

Question four involved determining if there was a statistically significant difference among the three groups of teachers with differing experience levels (novice, 2nd- and 3rd-year teachers, and teachers with 3 or more years of experience). As in questions 1-3, the experience levels of the teachers were based on reading teachers who served the students two years prior to test data collection. For example, in ANOVA – 1, comparing scores of second grade students, the novice scores were derived from students who were taught by a novice teacher in kindergarten two years earlier. Likewise, the 2nd- and 3rd-year scores were derived from students who were taught by a 2nd- or 3rd-year teacher in kindergarten two years earlier. Because the groups were classified on only one dependent variable (NCE scores), a one-way Analysis of Variance (ANOVA) was used to measure the differences. Three one-way ANOVA tests were performed, one on each grade level from which

test data were gathered. Table 19 illustrates the ANOVA design for question four.

In addition, the researcher chose to report effect size for three reasons.

> First, reporting effects facilitates subsequent meta-analyses incorporating a given report. Second, effect size reporting creates a literature in which subsequent researchers can more easily formulate more specific study expectations by integrating the effects reported in related prior studies. Third, and perhaps most importantly, interpreting the effect sizes in a given study facilitates the evaluation of how a study's results fits into existing literature, the explicit assessment of how similar or dissimilar results are across related studies, and potentially informs judgment regarding what study features contributed to similarities or differences in effects. (Thompson, In press)

Table 19

Independent Variable : Experience Level

Novice	In 2nd & 3rd year	In 4th or > year	One-way ANOVA
2nd Grade Reading NCE	2nd Grade Reading NCE	2nd Grade Reading NCE	ANOVA - 1
3rd Grade Reading NCE	3rd Grade Reading NCE	3rd Grade Reading NCE	ANOVA - 2
4th Grade Reading NCE	4th Grade Reading NCE	4th Grade Reading NCE	ANOVA - 3

The drawback to ANOVA, in this instance, was that specificity was lost. The resulting *F* from the analysis told there was a significant difference but not which groups were significantly from each other and which were not. Therefore, the Bonferroni corrected alpha was also included as a part of the ANOVA, J. Petrosko (personal communication, June 11, 2005). The data used for this analysis meets the five assumption of the ANOVA:

1. that the scale on which the dependent variable is measured has the properties of an equal interval scale
2. that the measures within each of the groups are independent of each other
3. that the source population(s) from which the samples of measures are drawn can be reasonably supposed to have a normal distribution
4. that the groups of measures have approximately equal variances
5. the differential effects of the conditions are consistent among the subjects (Gravetter & Wallnau, 2000)

The Statistical Package for the Social Sciences (SPSS) was used for analysis purposes because of its comprehensive and integrated capabilities in managing, analyzing, and displaying data (Gall, Borg, & Gall, 1996).

Assumptions and Limitations

Assumptions

In addition to the assumptions required for use of the single sample *t*-test and the one-way ANOVA, this study assumed that the reported scores from the Tennessee Department of Education were accurate and valid. The study also assumed the accuracy of the school system's management software for students and teachers.

Limitations

Three facts limited the generalizability of this study. The lack of racial and ethnic minority students included in this study limits its application to all settings. The study included only elementary teachers, limiting the knowledge gained to these grades and excluding secondary teachers and students. The study was also based on the reading NCE score limiting the transfer to mathematics and the sciences. The study was further limited by the noise of intervening years between the instruction and the data collection. The students in the study had experienced two teachers between primary reading instruction and data collection. While this noise might have interfered, overpowering the effect of the reading teacher, this would mean that the two intervening teachers had overpowered the effect of the earlier teacher.

CHAPTER IV

RESULTS

The purpose of this quantitative exploratory study was to investigate the residual or persistent differences among reading achievement of students taught by teachers with differing levels of experience. The researcher investigated the effect of teachers' years of teaching experience on student achievement two years later. Specifically, reading achievement scores of students of novice teachers, teachers with one to two years of experience, and teachers with three or more years of experience were examined two years after being taught in the primary grades by these teachers. Reading was chosen as the subject area for the study because most academic areas display some dependence on success in this skill.

Educational researchers and teachers have long recognized the importance of reading. Simply put, children who enjoy reading and frequently do so find greater success in school and in life. Making matters worse for students who do not enjoy reading, students who struggle with reading have historically received relatively poor instruction (Allington, 1994).

As stated in Chapter I, the study was designed to answer the following four questions:

1. Is there a statistically significant difference between the reading achievement, measured two years later, of students taught by a novice teacher in primary grades (kindergarten, one, and two) and the state mean for the appropriate grade level?
2. Is there a statistically significant difference between the reading achievement, measured two years later, of students taught reading by a second or third year teacher in primary grades (kindergarten, one, and two) and the state mean for the appropriate grade level?
3. Is there a statistically significant difference between the reading achievement, measured two years later, of students taught reading by a teacher with three or more years of teaching experience (teachers in their fourth or greater year of teaching) in primary grades (kindergarten, one, and two) and the state mean for the appropriate grade level?
4. Is there a statistically significant difference among the two-year-later reading achievement of the three groups of students in questions one, two, and three?

The independent variable for the statistical analyses of these questions, described in Chapter III, was the experience level of primary reading teachers. The dependent variable was the reading NCE score on the

TerraNova Achievement Test published by CTB/McGraw Hill and administered annually in the state of Tennessee. The results of these analyses are offered in this chapter.

The analyses results are presented in three sections. The first includes a description of the participants, including selection method, number of participants, and mean NCE scores for each group. The second section contains statistical analysis tables which illustrate the findings on the relationships between variables and interpretive comments on these relationships. In the third section, the findings on each of the research questions are summarized.

Description of the Participants

The study used archived longitudinal data from school years 1999-2000, 2000-2001, 2001-2002, 2002-2003, and 2003-2004 in a small suburban/rural school system in middle Tennessee. All 4,588 kindergarten, first, and second grade students in the school system during the first three years of this time (1999-2000, 2000-2001, and 2001-2002) were identified by purposeful sampling. Students in each of the identified three grade levels were grouped for analysis according to the experience level of their primary reading teacher (novice, one to two years' experience, and three or more years of experience).

This categorization yielded nine groups of students for comparison. The groups were (a) kindergarten novice, (b) kindergarten with one or two

years of experience, (c) kindergarten with three or more years of experience, (d) first grade novice, (e) first grade with one or two years of experience, (f) first grade with three or more years of experience, (g) second grade novice, (h) second grade with one or two years of experience, and (i) second grade with three or more years of experience.

Of the 4,588 students identified originally, two-year-later test data were available for 3,385 of these students. There were no test data on the other 1,203 students. Three possible explanations were: (a) absence during the week of testing (approximately 15), (b) retention (approximately 50), and (c) withdrawal from the system during the two years between identification and testing (approximately 1138). This system adjoins a large metropolitan system which may contribute to the high mobility rate.

In addition, the county has a high percentage of small "starter homes" and a low cost of living compared to other neighboring counties. These facts also contribute to the high mobility rate because as families become more affluent, they move to a more affluent area. When test-score lists for students were compiled by the researcher, no identified teacher's students were all still in the system two- years later. Each teacher was missing approximately five students in the two-year-later compilation. Two-year-later reading achievement data were collected on the 3,385 students who remained in the system and were tested two years later (2001-2002, 2002-2003, and 2003-2004).

Students of teachers on alternative or interim licenses were not included in the study. These teachers were not included as novice because they had not yet completed the initial licensure process. They were lacking either coursework or passing scores on PRAXIS tests. Most of them had been employed in this system with experience from other states, and were therefore experienced, but not licensed in Tennessee. Additionally, the students of teachers who serve exceptional populations solely (i.e., Title I), were not included. Other than these exclusions all student scores were included in the study.

Initially, 250 teachers were identified in the three grade levels used in the study. Nine of these did not meet the criteria for selection. Six taught self-contained Title I classrooms, and three taught on an interim license during their first year in the system. The exclusion of these teachers meant that students of 241 teachers participated in the study. The experience levels ranged from novice to 31 years. The mean experience level in the four or more years category was 12.06 years. In the experienced categories (one or two years and 3 or more years), not all of the teaching experience was always in the system studied. Many of the teachers had previous experience in other systems and states. This was not a consideration for elimination from participation in the study. Total years of experience was the only variable considered.

This decision to use the single variable of *teaching experience* was based on the findings of McGaffrey, Lockwood, Koretz, & Hamilton (2003), who contended there was no current method to "disentangle true teacher effects from student background characteristics in the presence of classroom-level variables and contextual effects and correlation between true teacher effects and student characteristics" (p. 74). In addition, the school populations are not greatly disparate in poverty levels and additional programs (as discussed in Chapter III). The schools serving students with higher poverty levels provide Title I services for these students. This service is the only outwardly discernible difference in the schools. With the findings of McGaffrey et al. in mind, and because this study was conducted in Tennessee, where Sanders(1996, 1997) studied teacher effects through TVAAS, the researcher elected to not control for student, school, and system variables. Table 11 in Chapter III provided details on each school, the teachers, and the student population.

Table 20 presents the number of students in each of the nine groups, organized by grade level. Students who withdrew from the system and those for whom there were no test scores (n = 1,203) were deleted from the study.

Table 20

Descriptive Analysis of Overall Student Sample

Independent Variable	Original *N*	*N* with no score	Final *N* in study
Kindergarten Novice	113	46	67
Kindergarten Year 1 or 2	211	72	139
Kindergarten Year 4 or more	1188	371	817
1st Grade Novice	112	34	78
1st Grade Year 1 or 2	232	61	171
1st Grade Year 4 or more	1214	325	889
2nd Grade Novice	157	44	113
2nd Grade Year 1 or 2	167	28	139
2nd Grade Year 4 or more	1194	222	972
Total	4588	1203	3385

Statistical Analysis

Based on a review of the literature on novice teachers and effective teachers, and the relationship of effective teachers on student achievement, null hypotheses were posited for all the tests conducted in the study. This section includes the results of each of the nine *t*-tests comparing teacher experience level to student achievement level two years later (questions 1-3). As outlined in Chapter III, effect size (*d*) was also calculated as [(Experimental group sample mean) – (Control group sample mean)] divided by (Control group SD). The SD for the TerraNova Achievement test is 21.

"By convention effect sizes are interpreted as follows: (a) .20 is small, (b) .50 is medium, and (c) .80 is large," J. Petrosko (personal communication, June 11, 2005). For interpretative purposes, effect sizes were calculated for each analysis in this study, which showed significant differences. The results were classified by the ranges recommended by Glass and Hopkins (1996).

The results of the three one-way ANOVA tests to determine variance among the three experience levels in each grade are also presented. The Tukey's Honestly Significant Difference (HSD) Test was performed post hoc when appropriate. The purpose of this calculation was to determine, specifically, the extent of the differences, if the null hypothesis was rejected.

Table 21 shows the reading achievement NCE scores for each experience level and grade in the study. One significant aspect of the scores is

that the majority of the nine identified groups are all above the state mean. This fact is meaningful when interpreting test results later in this chapter.

Table 21

Group Mean for Sample Groups Two Years Later

Independent Variable	Group Mean	State Mean
Kindergarten Novice	60	52
Kindergarten Year 1 or 2	62	52
Kindergarten Year 4 or more	60	52
1st Grade Novice	57	52
1st Grade Year 1 or 2	59	52
1st Grade Year 4 or more	56	52
2nd Grade Novice	57	52
2nd Grade Year 1 or 2	51	52
2nd Grade Year 4 or more	57	52

Table 22 summarizes the analysis performed for question one: Is there a statistically significant difference between the reading achievement, measured two years later, of students taught by a novice teacher in primary

grades (kindergarten, one, and two) and the state mean for the appropriate grade level?

The findings of two-tailed *t*-tests 1-3 comparing the results of the novice primary teachers' students (two years later) to the state reading achievement NCE mean of 52 are shown. For kindergarten students taught by novice teachers, the difference between the means is statistically significant $t(66) = 3.51$, $p < .0055$. The difference between the means for first grade students is not statistically significant at the .0055 $t(77) = 2.080$, $p > .0055$. The test results for this group of second grade students two years later is statistically different $t(112) = 2.838$, $p < .0055$.

Table 22

Student Scores of Novice Primary Teachers Two Years Later (*t*-tests 1-3) Compared to State Mean (52 NCE)

Grade Level	*N*	*M*	*SD*	*M* - 52.0	*df*	*t*	Sig. (2-tailed)
Kindergarten	67	60.27	19.27	8.27	66	3.512	.001
1st Grade	78	56.90	20.80	4.90	77	2.080	.041
2nd Grade	113	56.66	17.47	4.66	112	2.838	.005

Table 23 shows the findings of *t*-tests 4-6 for comparing the reading achievement results of the first and second year primary teachers' students two years later to the state mean NCE of 52 for question two:

> Is there a statistically significant difference between the reading achievement, measured two years later, of students taught reading by a second or third year teacher in primary grades (kindergarten, one, and two) and the state mean for the appropriate grade level?

The difference between the means for kindergarten students is significant at the .0055 level $t(138) = 7.025$, $p < .0055$. The difference between the means for first grade students is statistically significant $t(170) = 4.571$, $p < .0055$. For second grade students the analysis failed to reject the null hypothesis at the .0055 level $t(138) = -.784$, $p > .0055$.

Table 23

Student Scores of Second and Third Year Primary Teachers Two Years Later (*t*-tests 4-6) Compared to State Mean (52 NCE)

Grade Level	*N*	*M*	*SD*	*M* – 52.0	*df*	*t*	Sig. (2-tailed)
Kindergarten	139	61.99	16.77	9.99	138	7.025	.000
1st Grade	171	58.56	18.76	6.56	170	4.571	.000
2nd Grade	139	50.64	20.44	-1.36	138	-.784	.434

Table 24 shows the findings of *t*-tests 7-9 comparing the two-year-later reading NCE scores of students taught primary reading by a teacher with more than three years of experience, to the state mean of 52 for question two:

> Is there a statistically significant difference between the reading achievement, measured two years later, of students taught reading by a teacher with three or more years of teaching experience in primary grades and the state mean for the appropriate grade level?

In all three grade levels, the difference is statistically significant. The difference between the means for kindergarten students is statistically significant $t(816) = 11.859$, $p < .0055$. For first grade students, the difference between the means is statistically significant $t(888) = 5.139$, $p < .0055$. The null hypothesis was also rejected for the second grade students at the .0055 level $t(971) = 8.781$, $p < .0055$.

Table 24

Student Scores of Fourth or More Year Primary Teachers Two Years Later (*t*-test 7-9) Compared to State Mean (52 NCE)

Grade Level	*N*	*M*	*SD*	*M* – 52.0	*df*	*t*	Sig. (2-tailed)
Kindergarten	817	60.13	19.60	8.13	816	11.859	.000
1st Grade	889	55.42	19.87	3.42	888	5.139	.000
2nd Grade	972	57.01	17.79	5.01	971	8.781	.000

The scores in the preceding tables (22, 23, and 24) were reported by experience level because this was the format of the questions posed in Chapter III. The following tables (25, 26, and 27) report the same findings by grade level. Reporting the findings in this format allows the reader to see the differences, or lack of differences by grade level. The researcher calculated effect size on each *t*-test which showed an initial effect (Glass & Hopkins, 1996).

Table 25 illustrates the findings of *t*-tests 1, 4, and 7, comparing the results of the novice, second and third year, and fourth or more years teachers' kindergarten students (two years later) to the state mean of 52. For all levels of teacher experience, students significantly exceeded the state

mean of 52. Effect sizes for novice and 4th year or more were .39, between small to moderate. The effect size of 2nd- and 3rd-year was .48, moderate.

Table 25

Scores of Students with Kindergarten Teachers at Three Levels of Experience Two Years Later Compared to State Mean (52 NCE)

Experience Level of Kindergarten Teacher	*N*	*M*	*SD*	*M* – 52.0	Sig. (two-tailed)	*d*
Novice	67	60.27	19.27	8.27	.001	.39
2nd & 3rd Year	139	61.99	16.77	9.99	.000	.48
4th year or more	817	60.13	19.60	8.13	.000	.39

Table 26 illustrates the findings of *t*-tests 2, 5, and 8; comparing the results of the novice, second and third year, and fourth or more years first grade teachers' students to the state mean of 52, two years later. The researcher obtained *d* as the difference between the means, M_1 - M_2, divided by standard deviation, σ, of the of the NCE scores. Effect sizes for 2nd- and 3rd-year teachers were .31, between small to moderate. The effect size of 4th year or more was .16, small.

Table 26

Scores of Students with First-Grade Teachers at Three Levels of Experience Two Years Later Compared to State Mean (52 NCE)

Experience Level of First Grade Teachers	*N*	*M*	*SD*	*M* – 52.0	Sig. (2-tailed)	*d*
Novice	78	56.90	20.80	4.90	.041	–
2nd & 3rd Year	171	58.56	18.76	6.56	.000	.31
4th year or more	889	55.42	19.87	3.42	.000	.16

Table 27 illustrates the findings of *t*-tests 3, 6, and 9; comparing the results of the novice, second and third year, and fourth or more years teachers' second-grade students two years later to the state mean of 52. Effect sizes for novice and 4th year or more were .22 and .24, respectively, relatively small effects.

Table 27

Scores of Students with Second-Grade Teachers at Three Levels of Experience Two Years Later Compared to State Mean (52 NCE)

Experience Level of Second Grade Teachers	*N*	*M*	*SD*	*M* – 52.0	Sig. (2-tailed)	*d*
Novice	113	56.66	17.47	4.66	.005	.22
2nd & 3rd Year	139	50.64	20.44	1.36	.434	–
4th year or more	972	57.01	17.79	5.01	.000	.24

The study compared the three experience groups at each grade level to determine if there was a significant difference among the three groups of teachers with differing experience levels (novice, 2nd and 3rd year teachers, and teachers with 3 or more years of experience). The same data set was used for these tests which addressed Research Question Four. The researcher completed three one-way ANOVA tests to determine if there was a difference and a Tukey's Post Hoc to determine the difference. Results of the three tests are displayed in Tables 28, 29, and 30. In addition, the researcher calculated effect size where appropriate, using estimated omega squared ($\hat{\omega}^2$).

Table 28 compares the second grade scores of students taught by the three experience levels in kindergarten. The completed ANOVA indicated that the effect of years of teaching experience was not significant, $F(2, 1020 = .558, p > .57$. These data do not provide evidence of significant differences among the three levels of teaching experience.

Table 28

Analysis of Variance for Reading NCEs by Experience Level of Kindergarten Teacher (ANOVA – 1)

Source	*SS*	*df*	*MS*	*F*	*p*
Experience Level	412.36	2	206.18	.56	.57
Error	376821.88	1020	369.43		
Total	377234.24	1022			

Table 29 compares the third-grade scores of students taught by the three experience levels in first grade. The completed ANOVA indicated that the effect of years of teaching experience was not significant at the first-grade level, $F(2, 1135) = 1.89$, $p > .152$. Like the results for kindergarten, the results for first-grade teacher experience indicated no statistically significant difference.

Table 29

Analysis of Variance for First Grade Students' Reading NCEs by Experience Level (ANOVA – 2)

Source	*SS*	*df*	*MS*	*F*	*p*
Experience Level	1474.45	2	737.23	1.89	.15
Error	443554.53	1135	390.80		
Total	445028.98	1137			

Table 30 illustrates the third ANOVA completed involved the comparison of different experience level teachers at the second-grade level. For this test, the researcher used fourth-grade reading achievement scores. There was a statistically significant difference among the three groups of students, $F(2,1221) = 7.58$, $p < .05$. Therefore, the Tukey's HSD was performed to isolate and quantify the area(s) of difference. The result of this

post hoc test indicated that students (M= 50.64) of the one- to two-year group scored significantly lower than the other two groups (Novice M= 56.66; Three or more years M= 57.01). This was in keeping with Berliner's (1986) contention that the one- to two-year group matured at differing rates. Their scores were not as predictable as the other two groups. For this ANOVA, the effect size was 0.0107 (a small effect size).

Table 30

Analysis of Variance for Reading NCEs by Experience Level of Second Grade Teachers (ANOVA – 3)

Source	*SS*	*df*	*MS*	*F*	*p*	$\hat{\omega}^2$
Experience Level	4955.34	2	2477.67	7.58	.001	0.01
Error	399095.13	1221	326.86			
Total	404050.47					

Summary of Findings

The primary purpose of this quantitative exploratory study was to investigate the residual or persistent differences among reading achievement of students taught by differing levels of teaching experience. The researcher investigated the effect of teachers' years of teaching experience on student achievement two years later. Tables 8 and 10 in Chapter II illustrated that

there were differences in student reading achievement scores among the experience level groups, with scores improving as experience level increased. The questions in this study sought to determine if there was a persistent or residual relationship between teacher experience level in primary grades and reading achievement in later grades.

Research questions asked if there was empirical evidence to confirm differences in reading achievement that persisted over time among students who were taught by primary reading teachers with differing levels of teaching experience. Question one focused on the second-grade scores of kindergarten students who were taught by the three identified levels of experience. Question two investigated the third-grade scores of first-grade students who were taught by these same three levels. Question three explored the connection between second-grade students taught by the three levels of teachers and their reading achievement scores in the fourth grade. In each of the first three questions, these scores were compared to the state mean reading NCE (52).

The results of these three questions were similar but not identical. Table 31 lists the grades and experience levels included in the study and the decision on the null hypothesis for each group. Seven of the nine tests rejected the null hypotheses which stated that there would be no statistically significant difference between each group and the state mean of 52 NCE.

Table 31

Decisions on Null Hypothesis for each *t*-test

Independent Variable	Null Hypothesis Decision
Kindergarten Novice	Reject
Kindergarten Year 1 or 2	Reject
Kindergarten Year 4 or more	Reject
1st Grade Novice	Failed to reject
1st Grade Year 1 or 2	Reject
1st Grade Year 4 or more	Reject
2nd Grade Novice	Reject
2nd Grade Year 1 or 2	Failed to reject
2nd Grade Year 4 or more	Reject

Question four investigated the difference among the three experience level groups. The results of these three one-way ANOVAs to answer question four revealed that in seven of the nine comparisons, the tests failed to reject the null hypothesis. Table 32 lists the grades and experience levels included in the study and the decision on the null hypothesis for each group. The null hypothesis for question four stated that there would be no statistically significant differences among the experience level groups.

Table 32

Decisions on Null Hypotheses for each ANOVA

Independent Variables Comparisons	Null Hypothesis Decision	
Kindergarten Novice/ Year 1 or 2	Fail to reject	a
Kindergarten Novice/4 or more	Fail to reject	a
Kindergarten Year 1 or2/4 or more	Fail to reject	a
1st Grade Novice/Year 1 or 2	Fail to reject	b
1st Grade Novice/4 or more	Fail to reject	b
1st Grade Year 1 or 2/4 or more	Fail to reject	b
2nd Grade Novice/Year 1 or 2	Reject	c
2nd Grade Novice/4 or more	Fail to reject	c
2nd Grade Year 1 or 2/4 or more	Reject	c

Note: a- Overall ANOVA was not significant; b – Overall ANOVA was not significant; and c- Results of pairwise comparison following significant ANOVA

The findings from the analyses failed to reject the null hypotheses in most instances. These results indicate the lack of any relationship between student achievement and the experience level of their teachers in early grades.

Chapter V includes a discussion of the results of the study. Data are interpreted, and relationships between findings and theory are discussed. In addition, Chapter V provides the reader with implications and recommendations for further research.

CHAPTER V

SUMMARY AND CONCLUSIONS

Although the concept of effective teaching might seem straightforward and easily determined, the variables impacting effectiveness create a complex construct, which many researchers have attempted to describe and quantify. Over the past 50 years, researchers have described characteristics, behaviors, and effects of teacher expertise. While economists used the production function model (Greenwald, Hedges, & Laine, 1996; Hanushek, 1993), educational researchers explored the area of teacher expertise by describing effective strategies and actions characteristic of expert teachers (Brophy& Good, 1986; Marzano, Pickering, & Pollock, 2001).

Different studies addressed teacher effectiveness from a variety of perspectives, yet they are all in general agreement that effective teachers influence student achievement. They agree that the classroom teacher is a critical variable in today's classroom, with far-reaching influence (NCLB, 2001; Sanders, 1997; Stronge, 2002). Despite the differences in research findings, commonalities emerge that describe what an effective teacher knows and does.

In contrast, novice teachers struggle to acclimate themselves to a new environment which clashes with their beliefs systems, chokes their creativity, and overwhelms them with new responsibilities. During this survival stage (Berliner, 1986), they are faced with the same assignments and required to meet the same expectations that the more expert and more experienced teachers are given.

This study sought to determine if these equal but inequitable requirements for the novice have any long-term correlation to student achievement. The primary purpose of the investigation was to determine if the achievement differences between students taught by teachers at differing levels of teaching experience persisted into later grades. Was student achievement in later grades related to the experience level of earlier teachers?

A summary of the findings and a discussion are presented in Chapter V. In the context of the discussion, the researcher put forward both the limitations, cautions, and implication of the study results. This information is followed by recommendations for practitioners, suggestions for future research, and a conclusion to this dissertation.

Discussion of Research Findings

The research questions posed in Chapter I focused the study on the differences between novice and expert teachers and the subsequent effect of

these differences on student achievement. The first step in the investigation was a review of the pertinent literature on the subject.

Review of Literature

A thorough review of the literature included both quantitative and qualitative research in journals, technical reports, and books. The main topics were expertise in general, expert teachers, novice teachers, comparison of novice and expert teachers, and the effects of teacher expertise on student achievement. The synthesis of the research in these areas revealed a distinct difference between the characteristics, behaviors, and student achievement of novice and expert teachers. The reviewed literature left no doubt in the mind of the researcher that differences exist. The next part of the study involved determining if these differences were long lasting, affecting student achievement in the future.

Research Design

For the purpose of this study, all kindergarten, first- and second-grade students during a three-year period were selected by purposeful sampling. The identified students were categorized according to the teaching experience level of their teacher during the year of selection. The researcher then determined the achievement test reading NCE score for each student (who remained in the system) two years later. These scores were used for statistical analysis.

Statistical Analysis

Teacher experience levels were used as independent variables and student reading achievement test NCE scores composed the dependent variable. These were compared in a series of nine single sample *t*-tests and three one-way ANOVA tests. The *t*-tests used a Bonferroni corrected alpha. Effect sizes and Tukey's HSD post hoc tests were calculated for the ANOVA tests. These two procedures permitted the researcher to determine the actual differences between and among the groups.

Results of the *t*-tests show that there is a statistically significant difference between the mean score of most categories of teachers and the state mean. For eight of nine comparisons, school district means exceeded the state mean. The only exception to this is the group of students taught by a second- or third-year teacher in the second grade. Students in this group scored significantly lower than the state mean. This is contrary to the other findings and not readily explained. The researcher could discern no pattern in the scores. Novice teachers' students showed no pattern of mean differences higher or lower than the other two groups and neither of the other two groups produced mean differences generally higher or lower. The significant differences between the group means and the state mean were expected because all grade level scores were well-above the state mean (see Table 21).

A series of three one-way ANOVA tests compared the three group means for a particular instructional level (e.g., three different experience levels of kindergarten teachers). While there were differences in those groups (a plot of means was not linear), the differences were not sustained from novice, to one to two, to fourth or more. Student scores dropped dramatically in the one to two year group and rose sharply in the fourth or more year group. The resulting report of difference does not lend itself to conclusions. On the whole, there was no consistent statistically significant difference among the three experience level groups.

Limitations and Cautions

Results obtained from the analyses of student test scores related to the three identified levels of teaching experience are not generalizable to the general population of teachers. One major limitation to the above analyses is that it encompassed only six schools, all of similar size, background, and student population. Another limitation is the small sample of the two lowest experience level groups. The numbers in those groups were dramatically lower than in the highly experienced group. The researcher was limited to a very small group of novice teachers who may not have been representative of novice teachers as a whole. By limiting the study to primary grade teachers, the researcher limited the number of novices. In the system involved in the study, more experienced teachers tend to congregate in the primary grades. This was a major limitation.

In addition, the sampling approach was limited by resource and access constraints. The researcher did not have access to the large, fourteen-year state data set.

Implications

According to Bickman and Rog (1998), exploratory research is generally conducted to provide an orientation or familiarization with the topic under study, serving to enlighten the researcher about salient issues. Exploratory research helps focus future research on important variables. It is often a preliminary activity leading to a more rigorous descriptive and analytic study. Pedhazer (1997) contended that the purpose of data analysis is to "shed light on theory" (p. 8). Analysis of the data in this study endeavors to shed further light on the contentions of other researchers in the areas of novice teachers, teacher development, and student achievement.

This study was based on the theoretical concepts of Shulman (1986), Berliner (1986), and the work of Sanders (1998). Shulman (1986) theorized that effective teachers possess a special knowledge (pedagogical content knowledge). Berliner (1986) theorized that teachers progress through a series of development stages as they move from novice to expert. Sanders (1998) argued that the single most important factor affecting student achievement is the teacher. He contended that teacher effects are both additive and cumulative with little evidence of compensatory effects of more effective teachers in later grades. The residual effects of both very effective and

ineffective teaches were measurable two years later, regardless of the effectiveness of teachers in later grades.

The researcher selected one small aspect of these theories and studies in this research on novice and experienced teachers and their effects on student achievement. The study sought to contribute to the field of knowledge in these three areas of study and theory by methodically exploring teacher experience and its effect on student achievement.

Examination of the literature on novice teachers revealed that novice teachers did not share the attributes of effective teachers. Their planning, interactive teaching, reflections, knowledge base, classroom management, perceptions of student learning (difficulties), and instructional strategies do not compare favorably with those of effective teachers (Allen & Casbergue, 1997; Borko & Livingston, 1989; Cleary & Groer, 1994; Needels, 1991; Schempp, Tan, & Manross, 1998; Tochon & Munby, 1993; Westerman, 1991). With these apparent differences, where are the novice teachers in the continuum from ineffective to effective? This study contributed evidence to the investigation of the impact of teachers on student achievement and focused attention on a group which has not previously been quantified.

Previous research on novice teachers has been qualitative, usually focusing on specific subject matters (science and physical education), and with very small samples (usually one or two teachers). This study began the

process of studying a relatively large number of novice teachers and their impact on student achievement.

Recommendations for Practitioners

There was no persistent or residual effect of novice teachers. This study did not identify residual or persistent differences in the two-year-later scores of students taught primary reading by teachers with differing levels of experience. While the literature did not discuss the effect of novice teachers on their current or future students, it was specific in the differences in the teaching skills of novice and experienced teachers. Differences in content and pedagogical skills, coupled with the adjustments the novice must make to a new environment surely affect the learning of students.

While practitioners wait for more information on teacher effects, they must use the information available to them to implement and strengthen programs of teacher recruitment, induction training, placement, and retention.

Novice teachers enter the school environment ready to make a difference. School systems must provide scaffolding procedures that facilitate their growth from novice to expert. Mentoring programs, both formal and informal, give novices access to knowledge, advice, and encouragement that are not available to them in an isolated classroom.

DuFour and Eaker (1998) and Sparks (2005) speak to the community of learning within a school. "The quality of relationships among adults in

schools is a predictor of student learning, particularly in schools that are most challenged by the social ills of poverty and racism" (Sparks, p.91). He advocated teamwork as a part of every teacher's daily work.

Schools that follow the advice of these writers will provide avenues by which novice teachers can become experts. By establishing effective mentoring programs, encouraging professional learning communities, and planning for time in which teams can collaborate, effective administrators assist novices in the acclimation process from student to novice to expert.

Suggestions for Future Research

This study compared two-year-later student scores, by teacher experience level to the state mean, which each grade level exceeded. Future research should focus on comparing these groups to the local mean. This comparison would allow for a more specifically meaningful result to the practitioner. More sophisticated analysis, which screens for the confounding variables present in all classrooms, is advised. This small exploratory study did not control for such variables as grade retention, Title I services, socio-economics, race/ethnicity, or any teacher variables such as education level, education institution, or additional training received. Additionally, future studies that follow students over time, by examining progress annually, would contribute to the knowledge base on effective teachers.

Yet another possibility for future study would be a qualitative study, which ties instructional methodology to achievement. Teacher observations in

the classroom setting, coding strategies or behaviors, and associating these strategies and behaviors to student achievement would be helpful to the profession and the overall field of knowledge on effective teachers.

A fourth suggested study would involve identifying a much larger group of novice teachers, identifying student scores for that group to ascertain that they were truly below the system mean, and collecting two-year-later data for these students to measure for residual effect.

Most importantly, further and more sophisticated research is needed to compare the differing experience levels with the system mean. As increased attention to standards moves students closer to the desired goals of the federal and state agencies, it becomes more difficult to measure differences. This creates the need for more sophisticated research methods to measure this variable and others as they affect student achievement.

Conclusions

In this time of ever-increasing accountability for student achievement, the focus eventually falls on classroom teachers and their effectiveness. Teaching experience is but one variable in the larger array of variables at work in today's schools. However, it remains one of the most quantifiable. Research has demonstrated that it plays an important role in teacher expertise (Leinhardt & Greeno, 1986; O'Connor & Fish, 1998; Schempp, Manross, Tan, & Fincher, 1998; van Driel, Verloop, & de Vos, 1998). Without further research on exactly what role experience plays, we will continue to

miss an important piece to the education puzzle. Given this circumstance, it is clear that ongoing research is needed on how novice teachers gain expertise and the role they play in the total education of a student.

REFERENCES

Allen, R. M., & Casbergue, R. M. (1997). Evolution of novice through expert teachers' recall: Implications for effective reflection on practice. *Teaching and Teacher Education, 13*, 741-755.

Allington, R. L. (1994). Critical issues: What's special about special reading program for children who find learning to read difficult? *Journal of Reading Behavior*, 26, 95-115.

Alverman, D.E., & Boothby, P.R. "Children's Transfer of Graphic Organizer Instruction." *Reading Psychology*, 1986, pp. 7 (2), 87-100.

Anderson, T.H., and Armbruster, B.B. (1986) The value of taking notes during lectures, Technical Report No. 374. Cambridge, Massachusetts: Bolt, Beranek and Newman and Center for the Study of Reading, Urbana, Illinois, ED 277 996.

Appleton, K., & Kindt, I. (1999 March 28-31). *How do beginning elementary teachers cope with science: Development of pedagogical content knowledge in science.* Paper presented at the meeting of the National Association for Research in Science Teaching. Boston.

Beijaard, D., & Verloop, N. (1996). Assessing teachers' practical knowledge. *Studies in Educational Evaluation, 22*(3), 275-286.

Berliner, D.C. (1986, August/September). In pursuit of the expert pedagogue. *Educational Researcher, 15* (7), 5-13.

Berliner, D. (1988 Feb.). *The development of expertise in pedagogy.* Paper presented at the meeting of the American Association of Colleges for Teacher Education. New Orleans, LA.

Borko, H., & Livingston, C. (1989). Cognition and improvisation: Differences. in mathematics instruction by expert and novice teachers. *American Educational Research Journal, 26*, 473-498.

Bransford, J. D., Brown, A. L., & Cocking, R. R. (Eds.). (1999). *How people learn: Brain, mind, experience, and school.* Washington, D.C.: National Academy Press.

Bickman, L., & Rog, D. J. (1998). *Handbook of applied social research methods.* Thousand Oaks, CA: Sage.

Brickhouse, N., & Bodner, G. M. (1992). The beginning science teacher: Narratives of convictions and constraints. *Journal of Research in Science Teaching, 29,* 471-485.

Brock, B. L., & Grady, M. L. (2001). *From first-year to first-rate: Principals guiding beginning teachers* (2nd ed.). Thousand Oaks, CA: Corwin Press, Inc.

Brooks, J. G., & Brooks, M. G. (1993). *In search of understanding: The case for constructivist classrooms.* Alexandria, VA: Association for Supervision and Curriculum Development.

Brophy, J. E, & Good, T. (1986). Teacher behavior and student achievement. In M. C. Wittrock (Ed.), *Handbook of research on teaching* (3rd ed.). (pp.340-370) New York: Macmillan.

Bullough, R. V., Jr. (1987). First-year teaching: A case study. *Teachers College Record, 2,* 219-237.

Carlsen, W. S. (1999). Domains of teacher knowledge. In J. Gess-Newsome & N. G. Lederman (Eds.), *Examining pedagogical content knowledge* (pp. 133-144). Netherlands: Kluwer Academic Publishers.

Chase, W. G., & Simon, H. A. (1973). The mind's eye in chess. In W. G. Chase (Ed.), *Visual information processing.* (pp. 215-281). New York: Academic Press.

Chen, W., & Rovegno, I. (2000). Examination of expert and novice teachers' constructivist-oriented teaching practices using a movement approach to elementary physical education. *Research Quarterly for Exercise and Sport, 71,* 357-392. Retrieved October 28, 2001, from http://web3.infotrac.galegroup.com.

Chi, M. T. H., Feltovich, P. J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science, 11*(2), 121-152.

Cleary, M. J., & Groer, S. (1994). Inflight decisions of expert and novice health teachers. *Journal of School Health, 64*, 110-114.

Clotfelter, C. T., Ladd, H. F., & Vigdor, J. L. (2004). *Teacher sorting, teacher shopping, and the assessment of teacher effectiveness*, Manuscript in preparation, Duke University.

Cohen, D. K., & Hill, H. G. (2000). Instructional policy and classroom performance: The mathematics reform in California. *Teachers College Record, 102*, 294-343.

Coleman, J S., Campbell, E., Hobson, C., McPartland, J., Mood, A., Weinfeld, F., & York, R. (1966). *Equality of educational opportunity.* Washington, DC: U. S. Government Printing Office.

Coley, R. J., & Coleman, A. B. (2004, September). *The fourth-grade reading* classroom. Retrieved December 30, 04 from http://www.ets.ort/research/pic/reading.pdf.

CTB/McGraw Hill. (n.d.). *CTB/McGraw Hill.* Retrieved February 12, 05, from http://www.ctb.com.

Cunningham, J. L. (1995). Binet's contextual study of memory. *Psychological Reports. 77*(3), 955-961.

Darling-Hammond, L. (1995). Inequality and access to knowledge. In J.A. Banks & C.A. McGee Banks (Eds.), *Handbook of research on multicultural education* (pp. 465-483). New York: Simon & Schuster MacMillan.

Darling-Hammond, L. (1997, November). Doing what matters most: Investing in quality teaching. New York, NY: National Commission on Teaching and America's Future. ED 145-183.

Darling-Hammond, L. (2000). Teacher quality and student achievement. *Education Policy Analysis Archives*, 8(1), Retrieved September 16, 2001, from http://epaa.asu.edu.

Deci, E. (1971). Effects of externally mediated rewards on intrinsic motivation. *Journal of Personality and Social Psychology*, 18(1), 105-115.

deGroot, A. D. (1965). *Thought and choice in chess.* The Hague, Netherlands: Mouton.

DuFour, R., & Eaker, R. (1998). *Professional learning communities at work: Best practices for enhancing student achievement.* Bloomington, IN: National Educational Service.

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

Education Commission of the States. (1998). *Designing and implementing standards-based accountability systems.* Denver, CO.

Ericsson, K. A., & Charness, N. (1994). Expert performance; Its structure and acquisition. *American Psychologist, 49*(3), 725-747.

Federal Deposit Insurance Corporation. (2004). In *Report for county overview.* Retrieved December 28, 2004, from http://www2.fdic.gov

Feiman-Nemser, S. (2001). From preparation to practice: Designing a continuum to strengthen and sustain teaching. *Teachers College Record, 103*(6), 1013-1055.

Ferguson, R.F. (1991). Paying for public education: New evidence on how and why money matters. *Harvard Journal on Legislation, 28,* 465-498.

Field, K. (1979). *Teacher development: A study of the stages in the development of teachers.* Brookline, MA: Brookline Teacher Center.

Gall, M.D., Borg, W.R., & Gall, J.P. (1996). *Educational research* (6th ed.). USA: Longman.

Ganser, T. (1997). Similes for beginning teaching. *Kappa Delta Pi Record, 33*(3), 106-108.

Glaser, R., & Chi, M. T. H. (1988). Overview. In M. T. H. Chi, R. Glaser, & M. J. Farr (Eds.), *The nature of expertise.* (pp.15-27). Hillsdale, NJ: Erlbaum.

Glass, G. V., & Hopkins, K. D. (1996). *Statistical methods in education and psychology* (3rd ed.). Boston: Allyn and Bacon.

Gobet, F. (n.d.). In *Overview of research on expertise.* Retrieved July 16, 2004, from http://www.psychology.nottingham.ac.uk/staff/Fernand.Gobet/C82INT/Expertise.html.

Goetz, M. E., & Duffy, M. C. (2001). Assessment and accountability across the 50 states. In *Consortium for Policy Research in Education: Policy Briefs*. Retrieved March 3, 2005, from Graduate School of Education, University of Pennsylvania: http://www.gse.upenn.edu/cpre/.

Goldhaber, D. D., & Brewer, D. J. (1996, January 19). *Why don't schools and teachers seem to matter? Assessing the impact of unobservables on educational productivity.* Paper presented at the meeting of the Econometric Society. San Francisco.

Gravetter, F. J., & Wallnau, L. B. (2000). *Statistics for the behavioral sciences* (5th ed.). Belmont, CA: Wadsworth/Thomason Learning.

Greenwald, R., Hedges, L. V., & Laine, R. D. (1996). The effect of school resources on student achievement. *Review of Educational Research, 66*, 361-396.

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

Gwynn-Pauquette, C., & Tochon, F. V. (2003). Preservice teachers reflections on risk taking: The dynamics of practice and experience while experimenting with innovation during student teaching. *Alberta Journal of Educational Research, 49,* 163-180.

Hamilton County Department of Education. (n.d.). *Systemwide test data.* Retrieved June 21, 05, from http://www.hcde.org/test-data/default.htm.

Hanushek, E. A. (1986). The economics of schooling: Production and efficiency in public schools. *Journal of Economic Literature, 24,* 1141-1177.

Hanushek, E. A. (1989). Assessing the effects of school resources on student performance: An update. *Educational Evaluation and Policy Analysis, 19*(2), 141-164.

Hanushek, E.A. (1993). Can equity be separated from efficiency in school finance debates? In E.P.Hoffman (Ed.), *Essays on the Economics of Education*, (pp.35-73). Kalamazoo, MI: W.E. Upjohn Institute for Employment Research.

Hanushek, E.A. (1996). School resources and student performance. In G. Burtless (Ed.), *Does Money Matter? The effect of school resources on student achievement and adult success*, (pp. 43-73). Washington, DC: Brookings.

Harrington, H. (1994). Teaching and knowing. *Journal of Teacher Education, 45*(3), 190-198.

Hastie, P. A., & Vlaisavljevic, N. (1999). The relationship between subject-matter expertise and accountability in instructional tasks. *Journal of Teaching in Physical Education, 19,* 22-33.

Hedges, L. V., Laine, R. D., & Greenwald, R. (1994). Does money matter? A meta-analysis of studies of the effects of differential school inputs on student outcomes. *Educational Researcher, 23*(3), 5-14.

Hewson, M. and P.W. Hewson. (1983). Effect of instruction using students' prior knowledge and conceptual change strategies on science learning. *Journal of Research in Science Teaching 20(8)*: 731-743.

Hinkle, D. E., Wiersma, W., & Jurs, S. J. (2003). *Applied statistics for the behavioral sciences* (5th ed.). Boston: Houghton Mifflin.

Jencks, C., Smith. M. S., Ackland, H., Bae. J. J., Cohen, D., Grintlis, H., Heynes, B., & Michelson, S. (1972). *Inequality: A reassessment of the effects of family and schools in America.* New York: Basic Books.

Katz, L. (1972) Developmental stages of preschool teachers, *Elementary School Journal, 73(1).* 50-54.

Kintsch, W. (1979). On modeling comprehension. *Educational Psychologist, 14,* 3-14.

Kounin, J. S. (1970). *Discipline and group management in classrooms.* New York: Holt, Rinehart, & Winston.

Lasky, S. (in press). A sociocultural approach to understanding teacher identity, agency and professional vulnerability in a context of secondary school reform. *Teaching and Teacher Education.*

Leinhardt, G., & Greeno, J. G. (1986). The cognitive skill of teaching. *Journal of Educational Psychology, 78*(2), 75-95.

Lesgold, A., Rubinson, H., Feltovich, P., Glaser, R., Klopfer, D., & Wang, Y. (Eds.). (1988). *The nature of expertise.* Hillsdale, NJ: Erlbaum.

Magnusson, S., Krajcik, J., and Borko, H. (1999). Nature, sources, and development of pedagogical content knowledge for science teaching. In J. Gess-Newsome & N. G. Lederman (Eds.), *Examining Pedagogical Content Knowledge* (pp. 95-132). Netherlands: Kluwer Academic Publishers.

Manning, B. H., & Payne, B. D. (1996). Mental deliberations during teaching episodes: Novice teacher versus expert teacher. *Teacher Education Quarterly, 23*, 57-67.

Marzano, R. J., Pickering, D. J., & Pollock, J. E. (2001). *Classroom instruction that works.* Alexandria, VA: Association for Supervision and Curriculum Development.

McCaffrey, D. F., Lockwood, J. R., Koretz, D. M., & Hamilton, L. S. (2003). *Evaluating value-added models for teacher accountability* (MG-158). Santa Monica, CA: Rand Corporation.

McElroy, S. W., & Pai, K. (2003 November 6-8). *Do teacher experience and educational attainment matter for student achievement?: New evidence from Texas.* Paper presented at the meeting of the Association for Public Policy Analysis and Management. Washington, DC.

Meyers, B., Meyers, J., & Gelzheiser, L. (2002). Observing leadership roles in shared decision making: A preliminary analysis of three teams. *Journal of Educational and Psychological Consultation, 12*(4), 277-312.

Missouri Department of Elementary and Secondary Education. (n.d.). In *Reading-Assessment Instruments.* Retrieved January 3, 2004, from Missouri Department of Elementary and Secondary Education Web Site: http://www.dese.mo.gov.

Moir, E. (1990). Phases of first-year teaching. *New Teacher News, 2*(2), 6-7.

Morine-Dershimer, G. & Kent, T. (1999). Source of teachers' pedagogical knowledge. In J. Gess-Newsome & N. G. Lederman (Eds.), *Examining pedagogical content knowledge* (pp. 21-50). Netherlands: Kluwer Academic Publishers.

Munby, H., & Russell, T. (1994). The authority of experience in learning to teach: Messages from a physics methods class. *Journal of Teacher Education, 45*(2), 87-95.

National Commission on Teaching & America's Future (1996). *What matters most: Teaching for America's future.* New York: National Commission on Teaching & America's Future.

Needels, M. C. (1991). Comparison of student, first-year, and experienced teachers' interpretations of a first-grade lesson. *Teaching & Teacher Education, 7,* 269-278.

Nemser, S. (1983). Learning to teach. In L. Shulman & G. Sykes (Eds.), *Handbook of research and policy* (pp. 150-169). New York: Longman.

O'Connor, E. A., & Fish, M. C. (1998, April). *Differences in the classroom systems of expert and novice teachers.* Paper presented at the annual meeting of the American Educational Research Association, San Diego, CA.

Okpala, C. O., Smith, F., Jones, E., & Ellis, R. (2000). A clear link between school and teacher characteristics, student demographics, and student achievement. *Education, 120,* 487-496.

Pedhazer, E. J. (1997). *Multiple regression in behavioral research.* New York: Holt, Rinehart, & Winston.

Porter, A. C., & Brophy, J. (1988). Synthesis of research on good teaching: Insights from the work of the institute for research on teaching. *Educational Leadership, 45*(8), 74-85.

Provasnik, S. J., & Stearns, C. K. (2003 April 21-25). *Teacher quality and student educational attainment: Findings from the National Education Longitudinal Study (1988-2000).* Paper presented at the meeting of the American Educational Research Association, Chicago.

Ralph, E. G. (1999). Oral-questioning skills of novice teachers: . . . Any questions? *Journal of Instructional Psychology, 26,* 286-298. Retrieved January 20, 2004, from http://web3.infotrac.galegroup.com.

Reio, T. G., Jr. (in press). Emotions as a lens to explore teacher identity and change: A commentary. *Teaching and Teaching Education.*

Rivkin, S. G., Hanushek, E. A., & Kain, J. F. (2002, February). *Teachers, schools and academic achievement.* Retrieved June 22, 04 from http://wc.wustl.edu/workingpapers/Hanushek_Kain_Rivkin_1.pdf.

Rosenbloom, P. S., & Newell, A. (1986). The chunking of goal hierarchies: A generalized model of practice. In R. S. Michalski, J. Carbonell, & T. Mitchell (Eds.), *Machine Learning: An Artificial Intelligence Approach II.* Los Altos, California: Morgan Kaufman.

Rosenholtz, S. J. (1989). Workplace conditions that affect teacher quality and commitment: Implications for teacher induction programs. *The Elementary School Journal, 89*(4), 421-439.

Rovegno, I. C. (1992). Learning to teach in a field-based methods course: The development of pedagogical content knowledge. *Teaching and Teacher Education, 8,* 69-82.

Rowan, B., Correnti, R., & Miller, R. J. (2002). *What large-scale, survey research tells us about teacher effects on student achievement: Insights from the "Prospects" study of elementary schools* (CPRE-RR-051). Philadelphia, PA: CPRE Publications, Graduate School of Education, University of Pennsylvania. (ERIC Document Reproduction Service No. ED 477 466).

Rudner, L. M. & Schafer, W. D. (2001). Reliability. *ERIC Digest,* ED458213.

Rust, F. O. (1994). The first year of teaching: It's not what they expected. *Teaching and Teacher Education, 10,* 205-221.

Sabers, D. S., Cushing, K. S., & Berliner, D. C. (1991). Differences among teachers in a task characterized by simultaneity, multidimensionality, and immediacy. *American Educational Research Journal, 28,* 63-88.

Sanders, W. L., & Rivers, J. C. (1996, November). *Cumulative and residual effects of teachers on future student academic achievement* (R11-0435-02-001-97). Knoxville, TN: University of Tennessee Value-Added Research and Assessment Center.

Sanders, W.L., Saxton, A.M., & Horn, S.P. (1997). The Tennessee Value-Added Assessment System (TVAAS): A quantitative, outcomes-based approach to educational assessment. In J. Millman, (Ed.) *Grading teachers, grading schools,* (pp. 137-162). Thousand Oaks, CA: Corwin Press.

Schempp, P. G., Manross, D., Tan, S. K., & Fincher, M. D. (1998). Subject expertise and teachers' knowledge. *Journal of Teaching in Physical Education, 17,* 342-356.

Schempp, P. G., Tan, S., Manross, D., & Fincher, M. (1998). Differences in novice and competent teachers' knowledge. *Teachers and Teaching: theory and practice, 4*(1), 9-20.

Schmidt, M., & Knowles, J. G. (1995). Four women's stories of "failure" as beginning teachers. *Teaching & Teacher Education, 11*, 429-444.

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

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

Simon, H. A. (Ed.). (1979). *Models of thought.* New Haven, CT: Yale University Press.

Snow, C. E., Burns, M. S., & Griffin, P. (Eds.). (1998). *Preventing reading difficulties in young children.* National Research Council. Washington, DC: National Academy Press.

Sparks, D. (2005). *Leading for results: Transforming teaching, learning, and relationships in schools.* Thousand Oaks, CA: Corwin Press.

Stanulis, R.N., Campbell, P.E. & Hicks, J. (2002). Finding her own voice: A beginning teacher's story of learning to honour her own voice in teaching. International Journal of Educational Action Research, 10(1), 45-65.

State of Tennessee, (1996). Tennessee code annotated, Charlottesville, VA: Michie Law Publishers.

Stronge, J. H. (2002). *Qualities of effective teachers.* Alexandria, VA: Association for Supervision and Curriculum Development.

Tashakkori, A., & Teddlie, C. (1998). *Mixed methodology: Combining qualitative and quantitative approaches.* Thousand Oaks, CA: Sage.

Tennessee Department of Education. (n.d.). *How to interpret the report card.* Retrieved March 27, 2005 from Tennessee Department of Education Web Site: http://www.k12.state.tn.us/education.

Texas State Board for Educator Certification Panel on Novice Teacher Induction Support System (1998, November 6). *Final Report.* Austin, TX: Information and Support Center.

Thompson, B. (In press). *Foundations of behavioral statistics.* http://www.coe.tamu.edu/~bthompson/index.htm New York: Guilford.

Tochon, F., & Munby, H. (1993). Novice and expert teachers' time epistemology: A wave function from didactics to pedagogy. *Teacher & Teacher Education, 9,* 205-218.

United States Census Bureau. (2000). In *Tennessee Quickfacts.* Retrieved December 28, 2004, from http://quickfacts.census.gov/qfd/states/47/47021.html.

United States Department of Education. (n.d.). *ED.gov.* Retrieved August 12, 2004, from http://www.ed.gov/policy.

United States Department of Education. (n.d.). In *Reading First.* Retrieved January 2, 2004, from http://www.ed.gov/programs/readingfirst/faq.html.

van Driel, J. H., Verloop, N., & de Vos, W. (1998). Developing science teachers' pedagogical content knowledge. *Journal of Research in science Teaching, 35,* 673-695.

Veenman, S. (1984). Perceived problems of beginning teachers. *Review of Educational Research, 54*(2), 143-178.

Vinovskis, M., (1999). *History & Educational Policymaking.* New Haven, CT: Yale University Press.

Wang, M.C., Haertel, G.D. and Walberg, H.J. (1990). What influences learning? A content analysis of review literature, *Journal of Educational Research, 84,* 1, 30-43.

Ward, P., & O'Sullivan, M. (1998). Similarities and differences in pedagogy and content: 5 years later. *Journal of Teaching in Physical Education, 17,* 195-213.

Watts, H. 1980. *Starting out, moving on, running ahead or how the teachers' center can attend to stages in teachers' development.* Teachers' Centers Exchange Occasional Paper No. 8. San Francisco: Far West Laboratory for Educational Research and Development.

Wayne, A. J., & Youngs, P. (2003). Teacher characteristics and student achievement gains: A review. *Review of Educational Research, 73,* 89-122.

Webster, W.J., Mendro, R. (1997). The Dallas value-added accountability system. In J. Millman (Ed.), *Grading Teachers, Grading Schools: Is Student Achievement a Valid Evaluation Measure*? Thousand Oaks, CA: Corwin Press, Inc., pp. 81–99.

Webster, W., Mendro, R., Orsak, T., & Weerasinghe, D. (1998). An Application of Hierarchical Linear Modeling to the Estimation of School and Teacher Effects. Paper presented at the annual meeting of the American Educational Research Association, April 13–17, 1998, San Diego, CA.

Wenglinsky, H. (2002). How schools matter: The link between teacher classroom practices and student academic performance. *Education Policy Analysis Archives, 10*(12). Retrieved February 3, 2004, from http://epaa.asu.edu.

Westerman, D. A. (1991). Expert and novice teacher decision-making. *Journal of Teacher Education, 42,* 281-201.

Williams, J., & Williamson, K. M. (1996). University support for novice teachers: Successes, struggles, and limitations. *Education, 116,* 366-371.

Wilms, W.W., & Chapleau, R.R. (1999). The illusion of paying teachers for student performance. *Education Week,* 19(10), 48.

Wilson, S., Shulman, L., & Richert, A. (1987). 150 different ways of knowing: Representations of knowledge in teaching. In J. Calderhead (Ed.), *Exploring Teachers' Thinking* (pp. 104-124). London: Cassell.

Wright, S.P., Horn. S.P., & Sanders, W.L. (1997) Teacher and classroom context effects on student achievement: Implications for teacher evaluation. *Journal of Personnel Evaluation in Education,* 11(1), 57-67.

Zepeda, S. J., & Mayers, R. S. (2001). New kids on the block schedule: Beginning teachers face challenges. *High School Journal, 84*(84), 1-14. Retrieved January 20, 2004, from http://web3.infotrac.galegroup.com.

APPENDIXES

Page

A. Letter of Consent (Cheatham County Schools 215

B. Letter of Emption (University of Louisville) 216

C. Letter of Exemption (Western Kentucky University) 217

APPENDIX A

Letter of Consent (Cheatham County Schools

CHEATHAM COUNTY
Board of Education
102 Elizabeth Street
Ashland City, Tennessee 37015

Interim Director
Lynn E. Seifert
Phone: (615) 792-5664
Fax: (615) 792-2551

7/22/05

Mrs. Seifert,

I am working on my Ph.D. in the cooperative program with University of Louisville and Western Kentucky University. I am currently completing my dissertation on "The Residual Effect of Novice Primary Teachers on Reading Achievement Scores". To complete this study I am requesting permission to use system personnel records, student management records, and student achievement scores. At no time in this study will any personal information such as names or social security numbers be used. All records will be kept confidential at all stages of the study. Thank you in advance for your consideration.

Sincerely,

Connie Mayo
Elementary Instructional Supervisor

The district approves the study "The Residual Effect of Novice Primary Teachers on Reading Achievement Scores" (Study Number 400.05). You have permission to use the records mentioned in your request.

Lynn Seifert
Interim Director

7-22-05
Date

The school system does not discriminate on the basis of age, sex, race, color, creed, religion, national origin or handicap in the operation of its educational programs and activities including employment practices

APPENDIX B

Letter of Emption (University of Louisville)



UNIVERSITY of LOUISVILLE

dare to be great

HUMAN SUBJECTS PROTECTION
PROGRAM OFFICE

University of Louisville
MedCenter One, Suite 200
501 E. Broadway
Louisville, Kentucky 40202-1798

Office: 502-852-5188
Fax: 502-852-2164

July 28, 2005

Joseph Petrosko, Ph.D.
(Connie Fort Mayo)
ELFH Department



RE: 400.05

Dear Doctor Petrosko:

The above study has been received by the Human Subjects Protection Program Office. It has been determined by the chair of the Institutional Review Board that the study is exempt according to 45 CFR 46.101(b) 2 since the research involves use of educational tests (cognitive, diagnostic, aptitude, achievement), survey procedures, interview procedures, or observation of that is not exempt under paragraph (b)(2) of this section if:

(i) the human subjects are elected or appointed public officials or candidates for public office; or
(ii) federal statute(s) require(s) without exception that the confidentiality of the personally identifiable information will be maintained throughout the research and thereafter.

The study is exempt only if information that could identify subjects is not recorded.

The purpose of this study is to analyze the residual or persistent differences among reading achievement of students taught by differing levels of teaching experience.

Since this study has been found to be exempt, no additional reporting, such as submission of Progress Reports for continuation reviews, is needed. Best wishes for a successful study. Please send all inquires and electronic revised/requested items to our office email address at hsppofc@louisville.edu.

Sincerely,

Laura D. Clark, MD
Biomedical IRB Chair

LDC/cm

APPENDIX C

Letter of Exemption (Western Kentucky University)

Office of Sponsored Programs
270-745-4652
FAX: 270-745-4211
sponsored.programs@wku.edu

WESTERN KENTUCKY UNIVERSITY

The Spirit Makes the Master

Western Kentucky University
1906 College Heights Blvd. #11026
Bowling Green, KY 42101-1026

In future correspondence please refer to HS06-004, July 20, 2005

Connie Fort Mayo
1006 Interstate Circle
Cedar Hill, TN 37032

Dear Connie:

Your revision to your research project, "The Residual Effect of Novice Primary Teachers on Student Achievement" was reviewed by the HSRB and it has been determined that risks to subjects are: (1) minimized and reasonable; and that (2) research procedures are consistent with a sound research design and do not expose the subjects to unnecessary risk. Reviewers determined that: (1) benefits to subjects are considered along with the importance of the topic and that outcomes are reasonable; (2) selection of subjects is equitable: and (3) the purposes of the research and the research setting is amenable to subjects' welfare and producing desired outcomes; that indications of coercion or prejudice are absent, and that participation is clearly voluntary.

1. In addition. the IRB found that you need to orient participants as follows: (1) signed informed consent is not required from each human subject as the data will be collected from a secondary source (Cheatham County, Tennessee, Board of Education; (2) Provision is made for collecting, using and storing data in a manner that protects the safety and privacy of the subjects and the confidentiality of the data. (3) Appropriate safeguards are included to protect the rights and welfare of the subjects.

   This project is therefore approved at the Exempt Review Level until July 15, 2006.

2. Please note that the institution is not responsible for any actions regarding this protocol before approval. If you expand the project at a later date to use other instruments please re-apply. Copies of your request for human subjects review, your application, and this approval, are maintained in the Office of Sponsored Programs at the above address. Please report any changes to this approved protocol to this office. A Continuing Review protocol will be sent to you in the future to determine the status of the project.

Sincerely,

Sean Rubino, M.P.A.
Compliance Manager
Office of Sponsored Programs
Western Kentucky University

cc: HS file number Mayo HS06-004

CURRICULUM VITAE

NAME: Connie Fort Mayo

ADDRESS: 1037 Dorris Winters Road
Chapmansboro, TN 37035

DOB: Springfield, Tennessee – November 4, 1947

EDUCATION: and TRAINING

B.S. – Austin Peay State University
Major – Elementary Education
Concentration – Library Science
1965-1968

MA. Ed – Austin Peay State University
Major – Administration and Supervision
1969-1974

Ed.S. - Austin Peay State University
Major – Administration and Supervision
2000-2002

Working toward Ph.D. – University of Louisville/Western Kentucky University
Major – Educational Leadership
2002-2005

Project Wild, WET, and Learning Tree Trainer
McNellis Compression Planning Trained
COMP Classroom Organization and Management Trained
Thinking Maps Trainer
Write from the Beginning Trainer
SACS Facilitator Trained

AWARDS :

Tennessee Supervisors' Association – Juanita Henson Distinguished Service Award – 2003

Austin Peay State University – Outstanding Education

Graduate Student Award - 2002

Tennessee Environmental Education – Administrator of the Year – 1992

PROFESSIONAL SOCIETIES:

Phi Delta Kappa
Tennessee Supervisors' Association
National Staff Development Council
Mid-Cumberland Supervisors Study Council

INVITED PRESENTATIONS:

Differentiated Instruction – TN Department of Education – March, March, and April, 2005

PROFESSIONAL ORGANIZATION LEADERSHIP:

National Staff Development Host Committee (2005-2007)

For Nashville meeting in 2007

TN Teacher Evaluation Development Committee (2005)

TN Librarian Evaluation Development Committee (2005)

TN Supervisors' Assoc. President 2003

Title II, Title VI State Steering Committee 1995-2000

TN Supervisors' Assoc. Sec. (2 2001-2002)

Mid-Cumberland Supervisors' Sec. (1996-1998)

TN Staff Development Council, Bd. of Directors (1998-1999)

TN Federal Projects Staff Development Planning Team (1992, 1993)

Goals 2000 Grant Reader 2001 and 2002

SACS Peer Review Team Chair 2000, 2001, 2002, 2003, 2004 (Various schools and system

National Staff Development Council TN Planning Committee (1996)