CHAPTER 1

INTRODUCTION

The advent of the standards movement in the late 1980's paired with the subsequent focus on school and district accountability measures have made local, state, and national educators and politicians demand results, in the form of increased student achievement primarily in the areas of reading and mathematics. As a result, individual schools as well as school districts across the nation face a variety of sanctions including loss of money, accreditation status, and in some cases control of their school systems. The passage of the No Child Left Behind Act in 2001 (Public Law 107-110) has elevated this focus on student achievement and school district accountability to a national level.

One chief component of the No Child Left Behind Act is the establishment of benchmarks that measure student progress in reading. A portion of the reporting process includes benchmarks for subgroups of students (major racial and ethnic groups, economically disadvantaged students, English Language Learners and students with disabilities), so that by the year 2014, all students (all defined as 100%) will leave third grade reading on grade level. This focus on reading, especially for these subgroups of students, is driven at least in part by the gap in reading achievement between students living in poverty and middle-class students. That gap has not decreased at all in the last 10 years (Taylor & Pearson, 2001, p. 180). The 2003 *Nation's Report Card* reported that fourth grade progress in reading over the last 12 years (1992 – 2003), as measured by the National Assessment of Educational Progress (NAEP), has remained unchanged. In this same report, students eligible for free or reduced lunch were 32% more likely to score below the "basic" level of reading achievement standards set by NAEP as those students who did not qualify.

The National Academy of Science, in their report commissioned by the federal government *Preventing Reading Difficulties in Young Children* (2000), stated that a systematic, school-wide reorganization of reading programs, at both the classroom and school wide levels, is pivotal in improving reading achievement in low-performing schools. They recommend that this emphasis on program improvement should be the number one priority in educational research for improving reading instruction in the primary grades (Snow, Burns & Griffin, 1998). The Coalition for Evidence-Based Policy, also commissioned by the United States Department of Education, points to the numerous programs (which they label "interventions") in reading that claim to be "evidence-based" or "research-based" or developed using "research", yet the evidence used is either nonexistent or poorly designed (*Identifying and Implementing*, p. iii). Both of these reports cite the need for the field of education to become more like the fields of welfare policy and medicine by calling on educators to use "scientifically-based research to guide their evaluative decisions about which programs to implement" (p. iv).

Current Context

School administrators are confronted with "evaluation" on a daily basis. They evaluate teachers who in turn are evaluating students whose achievement results are in turn many times used to evaluate the administrators. In addition, school districts with schools that have chronically failed to meet minimum academic standards are evaluated, and face public sanctions (financial and political) when achievement goals are not met. For the first time this year, schools in Virginia have no accreditation safety nets and are either fully accredited or not, based on student performance on state achievement tests. This status will be reported to the public via not only newspapers, but also through letters sent to individual households within each school's attendance zone. In addition to this public evaluation of a school's academic program, the federal

government, through the No Child Left Behind Act legislation discussed earlier, will engage in a similar public report of individual schools as well as an entire school district's ability to meet adequate yearly progress toward the goals of the act.

The reality of evaluating and reporting, both internally and externally, has become somewhat of an educational enigma. Every administrator, teacher, and parent wants the best possible education for the children for whom they are responsible. Citizens want and in some cases demand the same. Most of the local, state, and federal assessment and accountability measures are grounded in reading achievement (and in some cases mathematics, and at the federal level with NCLB eventually in science as well). In addition, three recently released and highly influential publications (National Reading Panel, 2000; Learning First Alliance, 1998; Snow, Burns, & Griffin, 1998) that focused on the components of an effective reading program have all very publicly re-proclaimed the crucial need for all students to engage in rigorous literacy instruction in the areas of phonemic awareness, phonics, vocabulary, fluency and comprehension instruction so they successfully leave third grade reading on grade level and are able to successfully meet the complex needs of nonfiction text. In fact, the National Reading Panel Report (2000) was used as the research base for effective reading instruction in the No Child Left Behind legislation.

This high degree of accountability at all levels coupled with a renewed emphasis on research-based core reading program instruction have escalated the need for administrators to begin or refocus their efforts at monitoring their literacy programs. While standardized test results may be in many cases the bottom line for measuring students' reading achievement, they may not be an accurate or valid representation of a school or district's literacy program.

Purpose of the Study

The purpose of assessing the impact of any reading program is to determine the effectiveness of the program, use the results to improve instruction, and ultimately increase student achievement. To put it simply, this type of program assessment answers the question, "is this program any good?" or "does this program produce the desired result?". These questions and their subsequent answers are both short and powerful; however, before these questions can be asked and answered, an even simpler one must be raised: Is the reading program that a school or school division has invested time and money in putting into place being fully implemented? In short, can a division or school administrator make decisions based on the results of a summative reading program evaluation when they cannot be assured that the program is being fully implemented in the first place?

For example a superintendent might ask a research and testing department to analyze the impact of the "Success for All" reading program that the district has chosen for implementation in all underachieving Title I elementary schools. Their analysis of test data reveals that student achievement in reading over the last two years in almost 75% of the schools utilizing the program has remained stagnant or in a few cases has decreased. Are these dismal test scores a direct result of the implementation of this particular reading program? Or, shouldn't the degree of implementation of this particular program be determined first before assessing the effectiveness of the program? Farr, Pritchard, and Greene (1999) label these two very different questions as the categories of product evaluation and process assessment. Product program evaluation seeks to assess the value or worth of a particular program while process program assessment (used in this study) is focused on finding problems with a program's components as the program is being implemented and fix them before the program is formally evaluated

(p.168). In fact, Stufflebeam (1983) argues that a program's goals, design, and implementation should be the focus of any program assessment model.

To achieve the national goal of all students reading by grade three, educators and researchers alike have searched for programs and practices that would ensure student success. There is a great deal at stake for schools and school districts as a result of the identification and utilization of these programs and practices ranging from full accreditation to sanctions to privatization.

The purpose of this study is to provide division and school administrators with a means of identifying the degree of implementation of the components of their reading program through the development and testing of a formative reading program implementation assessment model. This study extends the research in literacy program assessment and evaluation by analyzing both the overall implementation of a specific primary reading program as well as specific independent variables as a means of program improvement. The conceptual development of this model utilizes a review of the literature on program evaluation in general, studies that assessed the impact of particular primary literacy programs, and the Norfolk Public Schools Elementary Reading Plan.

Research Questions

The broader concept of literacy program evaluation and more specifically using a formative or process program implementation model has not been fully developed in a recent context. As a result, this study is guided by the following research questions:

1. How can the implementation of a school or district reading program be assessed in a valid, reliable, and efficient way?

2. What accounts for the greatest variance in the implementation of a district's primary literacy program?

Definition of Terms

The Joint Committee on Standards for Educational Evaluation developed 30 program evaluation standards for use in assessing evaluations of educational programs (*Program Evaluation Standards,* 1994). To assist in developing these standards, the committee defined key terms and concepts. While this study is not a formal program evaluation, it is guided by several terms related to one. A program is defined as "educational activities that are provided on a continuing basis" such as a school district's reading program, a medical educational program, or a professional continuing education program. A program is "the object of evaluation" (p. 2). The opposite would be a project, activities provided for a defined period of time, like a three-day professional development series, or a two-year demonstration project. Evaluation is defined as "the systematic investigation of the worth or merit of an object "(p. 3). Dressel (1976) further limits this definition by defining evaluation as "a systematic and formal collection and interpretation of relevant information "which serves as a basis for rational judgment in decision making situations" (p. 9). Worthen and Sanders (1987) define evaluation as simply "the determination of a thing's value" (p. 22).

Using these definitions, the purpose of a program evaluation is to determine, through a systematic means, the value of a given program and thus make decisions about it. More specifically, this study uses a formative or process program assessment to "improve a system's function and quality" (Tuckman, 1985, p. 11). The purpose of a formative program implementation assessment is to take an internal look at a particular program with program improvement being the goal or aim (p. 151).

Theoretical Framework

Mills (1999) cautions educators about the over-reliance of using questionnaires and surveys (labeled perceptual data) to determine the effectiveness of any given program (p. 3). The formal evaluation of a program is many times set up as an afterthought, or is used to determine if a school is meeting externally determined criteria or needs (*Critical Issue*, p. 2). Furthermore, many administrators as well as policy makers believe that the only true measure of a program's worth is the end result, student achievement on standardized tests (p. 10). However, Farr and Pritchard (1989) counter that the specific merits of a worthwhile program "cannot be understood by just collecting test scores to determine how well students are doing" (p. 163). An analysis of programs must focus on things that can be changed, and thus make classroom instruction more effective instead of simply identifying program objectives, matching these given objectives with a given set of student test scores, and using the degree to which these two variables match as the sole determiner of whether a program is effective (p. 162-163). This study will be an application of their idea.

In the fall of 1999, a group of 27 principals, teachers, reading specialists, and central office administrators in Norfolk (VA) public schools met weekly for seven months to review and discuss recent research on primary reading development and instruction. The result of these sessions was the development of a research-based district reading program that included specific required daily instructional segments for kindergarten (Appendix A), first grade (Appendix B), and second grade (Appendix C). These required daily segments are focused around five categories of effective literacy instruction (Figure 1) and were designed to meet the goals of the program: strengthening primary reading instruction to ensure that all students will read on grade

level by the end of third grade and continuously moving students toward powerful literacy by expanding their comprehension skills.

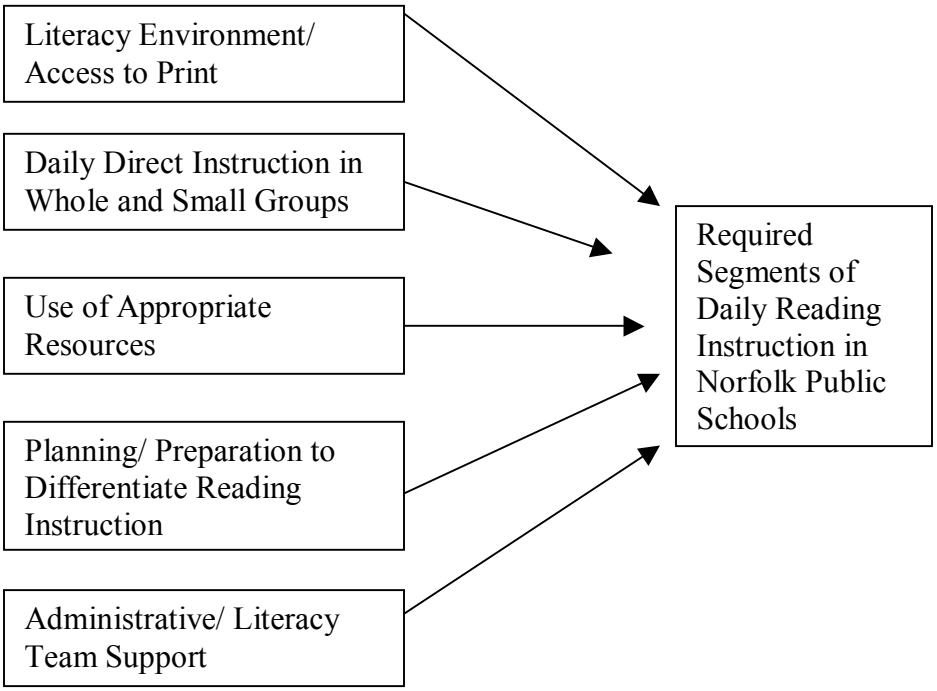

Figure 1. The five general domains of the Norfolk Public Schools Elementary Reading Plan

Theoretical Models

While this study is not a formal program evaluation, the idea of it in the field of education has been around for almost 100 years, beginning more formally in the 1920's. It was not until the Russian's launch of Sputnik in 1957 and the subsequent passing of the National Defense Education Act of 1958 (NDEA) where millions of federal dollars were poured into educational programs, primarily in math and science, that formal program evaluation methodologies began to be developed. The field was further expanded when the federal government passed the Elementary and Secondary Education Act of 1965 (ESEA), which included Title I (later Chapter I and later still Title I) legislation and funding. One component of ESEA was educational

research, development, and dissemination placing a strong emphasis on the replication of programs that had been deemed successful. In the following decades, this emphasis has continued and become even stronger as we have moved into the accountability era (Worthen & Sanders, 1987, pp. 16-19).

A variety of models have emerged and have been used over the past 50 – 60 years to assess the worth of educational programs. These approaches represent numerous conceptions about how an effective program evaluation should be structured (methodologies) and "metaphors an evaluator (could) use to understand the program under evaluation and the needs of the program stakeholders" (Fitzpatrick, 2004, p. 69). Worthen and Sanders (1987) note that at least 50 program assessment and evaluation models have been developed and used in the twenty years preceding their publication. Both Fitzpatrick (2004) and Gibton (2002) developed schemes to classify these different approaches. Fitzpatrick uses five classification schemas; Gibton uses three. Worthen and Sanders (1987) have developed six classifications while Smith (1994) uses seven descriptions as a means of classifying the various approaches to evaluation of programs. These classifications schemes are summarized in Table 1 according to type or label with a very brief description of the purpose or characteristics of each particular approach.

Although each of these approaches and models is unique and has value, the objectives-oriented and management-oriented approaches seem to be more closely aligned to the larger context of this study, and their applications are discussed in chapter two. An objectives-oriented evaluation specifies measured program objectives, uses objective instruments to gather data on the objectives and then identifies discrepancies between the objectives and the actual performance. The management-oriented approach also identifies program objectives, but evaluates the program at all stages of its development through a comprehensive look at

implementation as well as outcomes and metaevaluation (Fitzpatrick, 2004, p. 160). Formative, or interim evaluation, which will be used in this study, is rare because it is considered by many to be either too expensive, or too time consuming or both (Fink, 1995, p. 12).

This study is based on the development and testing of a primary reading program implementation instrument that is drawn from the district reading program's objectives. The study seeks to assess the implementation of the program in an overall scan to inform decision makers about improving the reading program. This study assesses the effectiveness of its implementation, but it does not assess the impact of its implementation on student achievement. Evaluation should be a process that provides useful information to decisions makers; the information that is provided should ultimately result in program improvement (Stufflebeam, 1983, p. 169).

Limitations

Just as there are a variety of models for determining the degree of program implementation as well as a program's effectiveness, there are a variety of problems associated with conducting a study of this nature and more specifically, conducting a study that includes collecting data about teaching. When any form of assessment or evaluation is used as the basis of decision-making and change, the results could be perceived as a way to resist change rather than a means to revise existing programs and thus encourage change. Others, who may not be clear about a specific program's goals, objectives, and methodologies might actually be clearer about these as a result of an assessment of it (Dressel, 1976, p. 25). Dunkin and Biddle (1976) noted the difficulty of collecting good examples of teaching as a part of any study. They cite the thousands of interactions that happen in rapid succession between students and a teacher during even the briefest of classroom observations. It is therefore very difficult to apply a rating scale

Table 1

Summary of Approaches to Educational Program Evaluation

Evaluation Label	Purpose	Proponent(s)
Objectives-oriented	To determine the extent to which specific program goals and objectives have been met at the conclusion of a program.	Fitzpatrick Gibton Smith Worthen and Sanders
Management-oriented	To determine the advantages and disadvantages of a particular set of decision alternatives throughout the development and implementation of a program.	Fitzpatrick Smith Stufflebeam Worthen and Sanders
Consumer-oriented	To provide information from an unbiased perspective on products or services that are perceived as being the same.	Fitzpatrick Worthen and Sanders
Expertise-oriented	To analyze the quality of a product or service using a professional	Fitzpatrick Worthen and Sanders
Participant-oriented	To develop data and conclusions from the perspective of the stakeholder	Fitzpatrick Gibton Smith Worthen and Sanders
Adversarial-oriented	To methodically incorporate opposing points of view into the evaluation of a program or service	Smith Worthen and Sanders
Policy-oriented	To inform local, state, or national policy makers	Smith

directly to a lesson. They also note the intrusive nature of a researcher observing in what is normally a "closed door" environment as well as the difficulty of using humans as subject in general (p. 26).

One problem with using humans (in this case, teachers) as subjects is the issue of representative or non-representative behavior. Flanders (1976) determined that unless the "within-group similarities are reasonably consistent and involve identifiable teaching patterns, the intended comparisons may be too weak to permit a fair test of possible differences in the educational program outcomes" (p. 383). For example, in theory a small number of teachers are selected as representative of the teaching population as a whole to allow the results of a study to be generalized to the population as a whole. Often the presence or absence of an observer forces a teacher to adjust his behavior as he is being observed and thus alters the results of the study. However, the problems with using humans as instruments can be compensated for in a variety of ways, including "triangulation (multiple observations or analyses) and refinements in the instrument. The shortcomings of human beings as instruments are more than compensated for by the quality and richness of the data they can gather" (Guba & Lincoln, 1981, p. 151).

Organization of the Study

Chapter one of this study provides an overview of current research and federal legislation that impacts early literacy instruction as well as a theoretical framework for this study and general conceptual models of program evaluation. Although this study is not a formal program evaluation, literature relative to its historical perspective and to the context of this study is provided. Chapter two examines research studies on attributes of effective primary reading programs that used qualitative and mixed methodologies. Chapter three focuses on the construction of a research-based model for assessing the fidelity of implementation of an

effective elementary reading program as well as a means of identifying those elements of the program that have the greatest variance in degree of implementation for future program improvement. Chapter four presents the results of classroom observations using the observation instrument developed for this study on the primary reading program implementation. Chapter five provides a summary of the results as well as a discussion of the limitations and future implications of implementing the model instrument based on this evaluation of it.

CHAPTER II

REVIEW OF THE LITERATURE

The purpose of this chapter is to review literature relative to the topic of assessing the implementation of primary reading programs. Several research studies including dissertations that have been conducted to evaluate the effectiveness of how school-wide programs in reading are analyzed and synthesized according to study type. The Educational Resource Information Center (ERIC), Education Full Text, Dissertation Abstracts International, and Ingenta databases were utilized for this review, using search terms that included reading with the following: program evaluation, educational evaluation, instruction, and evaluation as key words. In addition, the bibliographies from texts on program evaluation as well as the dissertations and studies selected for inclusion in this literature review were also searched.

All studies where the researchers use multiple sources of data (i.e. classroom observations, questionnaires, interviews, document analysis, test data) to draw their conclusions about the impact of primary reading programs were included. Experimental and quasi-experimental studies that rely on single sources of data (primarily student test scores in reading and the utilization of pre-test/post-test designs) as a means of determining characteristics of effective programs or general program effectiveness are excluded. Newmann, Smith, Allensworth, and Bryk (2001) note that “no studies have offered a thoughtful, systematic definition and exploration of school-level instructional program coherence” in any discipline. In fact, they found no study that addressed how instructional program coherence might increase student achievement or be an integral part of overall school improvement (p. 298).

Overview of the Studies

Purposes of the Studies

The focus of the studies that analyzed effective primary reading programs fall into three categories. Some studies examined the common characteristics of effective reading programs (Allington & Johnson, 2000; Guthrie, Schafer, Van Secker & Alban, 2000; Pressley, Rankin, & Yokoi, 1996; Pressley, Wharton-McDonald, Allington, Block, Morrow, et al. 2001; Taylor, Pearson, Clark, & Walpole, 1999; Taylor, Peterson, Rodriguez, & Pearson, 2002; Taylor, Peterson, Pearson, & Rodriguez, 2003; Walkup, 1997). Others attempted to identify the impact of a particular reading program on overall reading achievement (Barber, 2002; Bonifay, 1993; Curry & Zyskowski, 1999; Fisher & Adler, 1999; Fisher, Lewis, & Davis, 2002; Foertsch, 2000; Mehall, 2001; O'Neal, Snyder, & Spor, 2001). Two studies examined the impact of multiple programs on student achievement (Brown, Pressley, VanMeter, & Schuder, 1996; Mitchell & Wile, 2002).

Inclusion Criteria

All of the studies included in this review used a variety of qualitative methods including interviews, document analysis, questionnaires, and focus groups. Purely quantitative studies that used methods other than pretest/post test or posttest only designs were not found. Studies that used only student achievement information (quantitative data) as the single measure of a program's effectiveness are excluded as well. One researcher used a single site case study (Bonifay, 1993). Most of the studies used mixed methodology including surveys, interviews and questionnaires (Brown, Pressley, VanMeter, & Schuder, 1996; Curry & Zyskowski, 1999; Foertsh, 2000; Guthrie, Schafer, Van Secker & Alban, 2000; O'Neal, Snyder, & Spor, 2001; Pressley, Rankin, & Yokoi, 1996; Walkup, 1997). Some studies included pre-test/ post-test data

(Allington & Johnson, 2000; Fisher, Lewis, & Davis, 2002; Mitchell & Wile, 2002; Taylor, Peterson, Rodriguez & Pearson, 2002; Taylor, Peterson, Pearson, & Rodriguez, 2003; Taylor, Pearson, Clark, & Walpole, 1999) as one of several measures of a primary reading program's effectiveness.

The next two sections of Chapter II summarize the research studies that have been done relative to this study. The results are divided into two sections. The first section includes studies that attempted to isolate particular independent variables related to effective primary reading instruction as they impact student achievement in reading. Ten studies are included in this section. The following section includes studies that attempted to identify attributes of effective primary reading programs that exist in successful classrooms, schools, or school districts. Eight studies are included in the second section. In both sections, those studies where the researchers presented the most comprehensive picture of their particular study are expanded upon. Both sections of studies point to evidence that strong literacy programs have common instructional models that "guide the teaching and the learning climate…combining specific strategies and materials that inform the classroom environment" (Newmann, Smith, Allensworth, & Bryk, 2001, p. 299).

Results of the Studies

Impact of a Primary Reading Program on Student Achievement

The quality and level of implementation of a school district's reading/language arts program at the elementary and middle school levels was designed through the collection and analysis of multiple sources of data (Foertsch, 2000). These included gathering qualitative information about reading practices and instruction through classroom observations, a teacher questionnaire in closed response format (with a 57% return rate at the elementary level and a

65% return rate at the middle school level) to gather quantitative data on teachers' self-reported instructional strategies and areas of emphasis, group interviews for the purpose of gathering qualitative information concerning the successes and challenges of the current reading program, and individual teacher interviews with 40 staff members. These interviews were used to collect information on the staff members' specific views toward the reading program. Protocols were used to ensure consistency in the group and individual interviews and classroom observations. In addition, the researchers determined the percent of match between the district curriculum and the state standards and used state and national test scores to get quantitative information on student achievement in reading.

The study was a very thorough look at the quality and level of implementation of a single school district's reading/language arts program at the elementary and middle school levels using mixed methodology. For example, all reading, language arts, ESL, special education, and Title I teachers were observed for approximately 45 minutes at least once. Teachers at the grade levels where the state tests were administered were observed twice. Four focus groups, each lasting approximately three hours, were conducted with a mixture of teachers. The purpose of the focus groups was to gather information that could be used to inform the development of the individual teacher interviews. As this study evolved, the interview and classroom observation protocols were changed as recurring themes and patterns began to emerge.

Using reading achievement as well as interview and survey results, five areas of high consistency were isolated. These included a structured reading program, participation in staff development in best practices for reading instruction, use of planning time to analyze data and adjust their reading instruction, and a classroom focus that balanced comprehension and use of strategies with decoding in whole and small group settings. The use of new teaching approaches,

and the alteration of beliefs were all-essential in achieving an individual program goal, but individual teachers were determined to be implementing none, one, or more of these three dimensions.

Although the study presented a very extensive picture of student achievement using mean scores over time (seven years on the state reading tests; ten years on the national tests), there was no attempt to correlate this achievement with classroom practices. The researchers noted the low degree of curriculum alignment in their study design, yet no mention of the alignment (e.g. degree, impact on results) was made in their conclusions.

A very similar study was conducted in the Portland (OR) Public Schools (Mitchell & Wile, 2002). Surveys of 480 teachers and 22 principals on instructional practices, interviews with administrators and teachers, five classroom observations, and multiple choice reading test results were used in an attempt to determine the effectiveness of four reading intervention programs and professional development models as well as district adoptions of publisher reading materials currently in use in the district. Principals at schools making large gains in reading over the three years preceding this study recommended teachers who were interviewed and observed.

Multiple analyses of students' achievement on the state and district reading assessments were conducted. This included a longitudinal study of reading achievement in each of the schools using the various programs as well as a sample of comparison schools using cross tabulation. Second, by using intact cohort groups over three school years, the study correlated achievement outcomes with the various literacy programs, professional development models, and instructional strategies through the use of descriptive statistics and cross tabulations. In addition, reading assessment scores were used to select schools with large gains in the three years prior to the study to explore if their literacy program could be correlated with the achievement gains.

Finally, the relationship between the instructor's level of use of the reading programs and student achievement was investigated.

The qualitative analyses were limited to a summary of the priorities and needs for reading instruction that were identified by teachers and principals in interviews and surveys. Principals' comments from interviews and surveys were summarized. A determination of how teachers' needs and priorities differed by classroom characteristics, such as grade level, by examining mean scores on the questionnaires was also made.

In terms of student achievement, this study found no educationally significant differences in reading gains for students using the different reading programs in place in this district. Overall, reading achievement increased in this district, but in the schools with the largest gains, it does not appear to be the reading program that made the difference but rather the leadership of the principal and the size of the small instructional groups. This study also provided specific information about teacher reactions to the various programs and made recommendations for program improvement.

Teachers included as the subjects of this study were recommended by principals, so stratified random sampling procedures were not used. In addition, the majority of the results on effective instructional practices were obtained from a teacher questionnaire; therefore, many of the researchers' conclusions were based on self-reported data, based on how teachers indicated the frequency of use for each "best practices" strategy of activity that was listed. During this study, the lead evaluator took another position and subsequently had to abandon this study. As a result, not all of the evaluation questions and potential data that were originally proposed were addressed in this study. In addition, these restrictions could have narrowed the scope and direction of the research design. For example, through effect size statistics, the reading

achievement gains for students using each of the programs being evaluated and those in the comparison schools were not shown to be statistically significant or of practical importance. These results were used in lieu of more comprehensive statistical tests. This data, coupled with the questionnaire results that were similar to the achievement findings, were then used to make the conclusions.

Several other studies used a similar design and had similar results in determining the effectiveness of particular literacy programs. Curry and Ayskowski (1999) studied the impact of a specific summer reading program that used a balanced-literacy approach on increasing teachers' pedagogical knowledge as well as student achievement. Through the use of teacher and principal interviews, classroom observations, as well as student demographic and achievement information, the balanced approach to literacy instruction had a positive effective on student achievement as well as teacher utilization of the strategies. O' Neal, Snyder, and Spor (2001) looked at the relationship between the attributes of a state program initiative and the overall success of the program. This study, like the others that were focused on in greater detail in this section of the review, identified a helpful reading specialist and a principal as champions as more crucial than the attributes of the specific program through the integration of information recorded by three interviewers that established categories to reflect the nature of the responses for each interview question. The same state program initiative was used by Barber (2002) to determine if the program had any influence on the culture of schools. The major categories that emerged from her interviews included a correlation developed through interview and focus group transcripts between reading success and improved school culture and teacher efficacy, the program's impact on pedagogical changes in instructional design, and the positive impact of additional time on program implementation.

Field notes and documents were analyzed utilizing qualitative research methods to determine the impact of the state mandated development of literacy outcomes in a study conducted by Mehall (2001). The study was organized into three phases including the identification of schools that had identified reading as a goal for improvement, six days of fieldwork to conduct interviews, collect and examine documents, and make classroom observations, and a summary of the findings from the schools to determine if involvement in this program resulted in any significant changes in reading achievement. The results of the study included the acknowledgment that the state requirements had an impact on the school cultures, which in turn led to improvement in reading. The reading specialist was identified as the person who has the greatest impact on changes in reading instruction at the individual school level while Walkup (1997) and Mitchell and Wile (2002) both identified the principal as the one who has the significant characteristic in their studies. Mehall's study was limited by the time of year her study was conducted. Although schools were identified in the spring, the actual fieldwork was conducted during summer school. And, although the summer school program was very similar to the program conducted during the school year and was staffed by many of the same teachers, the attitudes and climate of a summer school environment could have had an impact on her findings.

A similar study was designed to determine which aspects of reading instruction had the greatest effects of growth in student reading achievement in nine high poverty schools representing rural and urban areas across the United States (Taylor, Peterson, Pearson, and Rodriguez, 2003). Two teachers per grade level (in grades 1 – 4) were randomly selected to be a part of the classroom observations and interviews. A total of 88 teachers and 792 students across these grades participated in the study. Stratified random sampling was used to select the students by asking participating teachers to divide their classes into lists of high, medium, and low

reading abilities. From these lists, three students from each group were selected as the sources for student achievement results.

Teachers who participated in the data collection were interviewed three times (fall, winter, and spring) for approximately 30 minutes each. Each teacher was also observed teaching reading three times for one hour per observation. A classroom observation system was used that combined note taking with a method of quantitative coding. Observers were trained by one of the research study members and had to demonstrate greater than 80% agreement in order to continue conducting classroom observations.

Using hierarchical linear modeling (HLM), the relationship between student's growth in reading achievement and teacher's practices during reading instruction were analyzed. Five attributes that contributed to students' growth in reading were identified and included amount of time spent on reading and reading instruction, coaching and involving students in active reading, higher level questioning, more small group than whole group settings, and comprehension skill and strategy instruction. Case study descriptions that illustrated each of these characteristics were also included as a part of the study.

Bonifay (1993) and Fisher, Lewis, and Davis (2002) both conducted similar studies of very prescriptive reading programs using both qualitative (case studies) and quantitative (pre-test/ post-test) measures. Bonifay found positive short-term and long-term impact as a result of the study. Teachers implemented the program correctly, and the students benefited as evidenced by both their reporting of strategy use during interviews, and their increased test scores. On the other hand, Fisher et al. determined just the opposite. In their study, they found uneven program implementation. Contrary to the expectations of both the program and the researchers, using cross tabulation and chi square analyses, there were no significant differences between children

in classes of different sizes, number of instructional groups in the class, nor the age position of a particular group in a multi-age class.

The previous section of this review summarizes study findings related to the relationships between the implementation characteristics of primary reading programs and subsequent increases in staff effectiveness and student achievement. All but one of the studies identified a series of attributes that result from the implementation of the given program. Each of the studies also had limitations. The primary limitation was in the way the teachers who were involved (through either questionnaire, interview or classroom observation) were chosen. Most were selected by a school staff member (either the principal or a reading specialist in the building) and not by the researchers. In other cases, the inclusion criterion for the selection of participants was not stated. In addition, most of these studies gathered information on teachers and their classrooms in limited geographic regions or in schools with homogeneous demographics.

The remaining studies in this review are included because each attempts to isolate and identify variables that characterize an effective primary reading program. While these studies are program neutral, their findings mirror many of those found in the program-specific ones that are summarized in the previous section.

Studies That Identify Effective Primary Reading Program Characteristics

The early reading programs in five high-performing, high-poverty schools were examined in a study that consisted of six descriptive cases and a cross-case analysis examining early reading programs in these schools (Fisher and Adler, 1999). The primary focus of each case study was a description of how the school's early reading program had been organized, developed, and sustained. Through individual case studies and in a cross-case analysis, patterns of how specific resources were allocated, between-class arrangements, and school-wide

structures that contributed to higher levels of student reading achievement emerged. Potential schools were identified through an examination of at least three years of state testing data in schools where at least half of the students were eligible for free or reduced lunch. Two schools were chosen from each of three states and were invited to participate in the study.

Structured interviews of 35 teachers and their respective six principals were conducted as well as approximately 12 hours of classroom observations per school. Informal conversation notes as well as selected artifacts from the reading programs were collected. From this information as well as the survey responses and selected public domain information about the schools, five key elements were identified that were responsible for the success of the students in these schools. They included a focus on student learning outcomes, multiple reading programs in every classroom, shared responsibility for student success, a print-rich classroom, strong leadership at the school level, and a veteran, knowledgeable staff.

The researchers acknowledged that the characteristics are an "appropriate starting place…but were not yet sufficiently well understood to generate a compelling theory or framework" (p. 5). For example, there was no attempt to determine the impact of the findings across the various factors that were identified as being effective. What were the relationships among and between the school-wide and other factors that promoted early reading achievement in these schools? The researchers summarized their raw data describing an early reading program in more conceptual terms (i.e. "shared responsibility", "strong leadership"). These generalized results may be difficult to replicate and implement at an individual school level.

In their study, Allington and Johnston (2000) used interviews and 10 full day observations of 30 fourth-grade teachers that were identified as exemplary in 24 schools in five states. Additional data were gathered in two semi-structured interviews with each teacher who

was observed, in interviews with students from each classroom, in the collection of classroom documents (e.g. writing samples and reading logs), and in end-of-year achievement test performance.

A post hoc analysis of achievement test gains in these classrooms was conducted resulting in the identification of teachers who produced greater than expected levels of literacy growth. The common features of the classrooms, which included collaborative conversations between teachers, use of multiple types of curriculum materials, a school-wide emphasis on literacy and an on-going evaluation of student work that focused on improvement, progress, and effort were identified. These were grouped in "common clusters". What the researchers did not do is attempt to move beyond these clusters and identify, for example, the sources and implications of classroom talk. While test data was used to validate that the identified teachers were indeed "successful", there was no indication of how the data was used except to indicate that (they are) "not satisfied with narrow notions of educational achievement such as those indicated by standardized test scores" (p. 24).

Another study examined a state mandated school reform (the development of learning outcomes in reading) in Illinois as a dependent variable to identify and analyze any changes in instruction and assessment (Walkup, 1997). This study also looked at the relationship between collaborative activities in a school and changes in classroom reading instruction and assessment through the use of a survey instrument designed specifically for the study that would yield both quantitative and qualitative information. The sample was limited to teachers of reading at grades three and four who were selected from a random sample of those schools that participated in a quality review process four years earlier. This extremely limited sample was a major problem for the researcher and possibly narrowed and even hindered her results.

The findings of the study indicated that teachers were knowledgeable about the state requirements, but did not believe that these requirements would impact reading instruction and assessment. The survey also revealed the belief that the principal's involvement, support, and provision of resources were most important to developing changes. Changes in reading instruction and assessment practices occurred that reflect the view of literacy as cognitive, constructive, and interactive including an emphasis on small group settings and an emphasis on planning instructional episodes that better met individual student needs.

A survey was used by Inman, Marlow and Barron (2004) to determine the effectiveness and effective features of a supplemental reading program. In their study, over 600 teachers in 153 different schools were surveyed to determine their perceptions of the various elements of a new reading program. In addition student performance information was gathered and compared for the year prior to the implementation of the program with the year following the implementation of it. Teachers who completed the survey reported strong agreement with the alignment of the new materials with their state standards and also strongly agreed that the materials were well organized which contributed to their successful implementation. In all but the sixth grade, student achievement scores rose as well.

Taylor, Pearson, Clark and Walpole (1996) conducted observations of 70 teachers in first through third grades from four different states for one hour each month for five consecutive months. They also used teacher interviews, a questionnaire, and teacher-kept logs from a subset of these teachers. Interviews with the 14 building principals involved in the study were conducted. A limited amount of student performance data was collected (4 students in each classroom observed). A composite score based on the overall school mean for these students' gain on individually administered reading measures and the school's average grade 3

standardized test score in reading was created and used to classify the schools involved in this study as most, moderately or least effective schools. Based on a summary of this information, by determining the percent of teachers who exhibited certain characteristics in each of the three school categories, seven variables were identified that characterize effective primary teachers in the "most effective" schools. These included the time spent in small or whole group reading instruction, the amount of time student spent in independent reading, and the way the teacher approached word recognition and comprehension instruction in a whole group setting.

This same lead researcher (Taylor, Pressley, & Pearson, 2002) in a later study used teachers from eight schools with high rates of poverty and interviews with principals to document effective program characteristics and classroom practices as well as classroom observations and test data. The purpose of the study was to describe the common teacher practices that were observed in classrooms and examine the relationships between these practices and students' growth in reading achievement. This study was broader in its sample. It included a more diverse set of schools and encompassed a broader range of grade levels (first through sixth grades). The later study's findings at the classroom level suggested that it is important to consider not only what teachers teach, but also how they teach when determining how to change reading instruction and increase reading achievement.

Thirty-three schools that demonstrated increased achievement for two successive years on a statewide performance assessment were identified and compared with schools that were decreasing in achievement in a study conducted by Guthrie, Schafer, VonSecker, and Alban (2000). Through the use of the instructional characteristics of the school reading program, 545 teachers completed a questionnaire on reading instruction that was constructed using research on classroom variables expected for effective reading comprehension. The response format was a

Likert-type scale. A preliminary study using the questionnaire with 225 teachers in 15 elementary schools for construct and content validity was also conducted.

The measure of student achievement in this study was the School Performance Assessment Program (MSPAP) used statewide in Maryland. The goal of this study was to identify characteristics of a school reading program that could be associated significantly with change in student achievement over time. Using a meta-analysis of six instructional characteristics (basal emphasis, books and resources, comprehension instruction, integrated reading and content, social collaboration, and writing) representing the school reading program to predict achievement change in reading and other content areas, the researchers compiled a profile of effective characteristics. This included an examination of the relationship between the individual instructional characteristics as compared to the variance in effect size for reading achievement change across each of the 33 schools that participated in this study.

The analytic unit in this study was the school and both the control and the experimental treatments were the achievement test given to independent groups of students. An analysis was used to test for the specific contribution of each of the six instructional characteristics as compared to the variance in effect size for reaching achievement change in each of the 33 schools involved in this study. The results indicated the positive effect of using but deemphasizing basal instruction, integrating reading instruction with other content areas, and using a wide variety of classroom materials for instruction. Less effective classrooms were characterized by the opposite – strong reliance on the basal, little integration of reading with other content areas, and sparse use of other resources.

This study was conducted in three school districts where most of the teachers were White, and well educated. Urban schools were not included in the study (none volunteered).

Because of this small, homogenous sample size, it was not reasonable to generalize the results of this study to an entire population of teachers. In addition, this study does not focus on effective instruction in the primary grades, so at best the results could be generalized and applied to upper elementary school classrooms and teachers.

Using classroom observations and interviews, Pressley, Wharton-McDonald, Allington, Block, Morrow, et al. (2001) identified teaching behaviors and characteristics that distinguished the most effective first grade teachers from the least effective teachers. For this study, 30 first-grade classrooms from five different states were used. At each participating school district or school, local personnel identified one outstanding teacher who promoted student literacy achievement (in the opinion of the school official) as well as one teacher who was perceived as being more typical but who taught students very similar to those taught by the outstanding teacher. After conducting extensive classroom observations and interviews, 11 behaviors that distinguished the most effective teachers were identified through the use of qualitative triangulation of the observational and interview data to make general conclusions about every teacher that was observed. These conclusions included the planning for and use of explicit teaching in response to individual student needs, encouragement of student self-regulation, reading skills and strategy instruction, and the scaffolding of process writing instruction coupled with high demands. Although the 30 teachers who participated in this study were observed repeatedly for documentation of their teaching and for observation of the reading and writing of students in their classes, the findings from only a third of the teachers (10) were used in the results. The study did not include an explanation of why the other 20 teachers were excluded.

In an earlier study Pressley, Rankin and Yokoi (1996) surveyed 83 primary (K – 2) teachers who were nominated by their supervisors as effective teachers of reading and writing.

The teachers responded to two questionnaires about their practices. Several of the self-reported practices were similar to those observed in the study done by Pressley et al. (2001). The strongest links were in direct instruction of reading skills and strategies, extensive reading, literature-rich classroom environments, and qualitatively similar instruction for students of all abilities. The study was limited to a survey that provided information about many elements of instruction, but did not give much insight into teachers' unique implementations of these instructional elements.

Nine schools (six that were successful; three that were not) representing urban, suburban, and rural areas in Vermont were identified by Mosenthal, Lipson, Russ, and Sortino (2003) in a qualitative study to determine the similarities and differences in approaches to early reading instruction. Interviews of 77 different teachers as well as building principals, curriculum coordinators, librarians, and district superintendents were conducted. The research team also made three separate classroom observations (in the fall, winter and spring) to each of the teachers that was interviewed as part of the study. Following the initial interview and three observations, a follow-up interview with each teacher was also conducted.

The study produced no common or consistent list of traits that were employed through either the successful classrooms or schools. Several common factors that were associated with high literacy achievement were identified. These included a significant amount of time that was devoted to students' reading. Each successful classroom had an extensive set of books. Time spent on instruction and practice in reading in these schools was significantly higher than the amount of time spent in the less successful schools. In addition, the successful buildings had a commitment to literacy and its improvement at all levels—superintendent, principals, reading specialists, and teachers.

The data from the less successful schools was used by the team to validate the information that was made about the successful schools. All of the characteristics of effective literacy instruction that were present in the successful schools were either not present or were less common in the ineffective ones. For example while in the successful schools there was a common vision or focus on literacy achievement, the less successful schools lack a common vision. The less successful schools had fewer books in classrooms and libraries than the successful ones.

Summary of Results of the Studies

These studies of primary literacy programs used an objective-oriented or a management-oriented approach to program evaluation by seeking to determine whether the goals of a reading program were being met (through increased student achievement in reading) or whether specific characteristics of an effective primary reading program were present. Most of the studies attempted to isolate particular variables (teaching characteristics or general characteristics) that could be attributed to the general reading success in classrooms, schools, or clusters of schools. In addition, these studies tied these common attributes to an independent variable that was a measure of literacy program "success" whether that was through inclusion in a teacher's classroom who had a history of good test scores or by simple identification by a building administrator or other literacy leader that the teacher was "good". In other studies, the researchers first identified key reading program elements, and then identified the presence or absence of those particular elements in previously designated classrooms or in schools or classrooms that had already demonstrated reading success. None of these studies used experimental research where random assignment of individuals (either students or teachers) was conducted – which may be impossible in the educational field because of ethical constraints. The

results of these studies can also be classified according to the domains of effective reading instruction that were used in this particular study (Table 2).

Table 2

Research Studies Organized by Literacy Study Domains

Domain of Effective Primary Literacy Instruction	Related Research Studies	
Environment/Print – a print rich classroom that includes a classroom library, word wall, and well defined areas for whole and small group instruction.	Allington, 2000 Fisher, 1999 Guthrie, 2000	Musenthal, 2003 Pressley, 1996
Daily Direct Instruction in Reading – time spent in whole and small groups developing reading skills and strategies.	Barber, 2002 Curry, 1999 Foertsh, 2000 Guthrie, 2000 Mitchell, 2002	Pressley, 1996 Pressley, 2001 Pressley, 2003 Taylor, 1996
Resources – Appropriate and a variety of resource materials to deliver reading instruction.	Fisher, 1999 Musenthal, 2003	
Planning/Preparation – Teachers develop plans that meet the literacy needs of all students.	Foertsh, 2000 Pressley, 1996 Pressley, 2001	
Administrative Support – Principal and literacy leaders promote, support and champion reading instruction .	Fisher, 1999 Mehall, 2001 Mitchell, 2002	Musenthal, 2003 O'Neal, 2001 Walkup, 1997

Summary of Chapter II

The literature, as it relates to the topic of effective primary reading programs was reviewed in Chapter II. The review of the literature was subdivided into two major headings: (a) studies that identified specific reading program attributes as they related to increased student achievement, and (b) studies that attempted to isolate variables of effective reading programs in schools or classrooms. In addition, each of the studies used in the review was classified according to the particular domain of effective reading instruction that was identified in the study's results or conclusions.

CHAPTER III

METHODOLOGY

This chapter describes the population and sample that was used in this study to assess the implementation of a primary reading program. The development, administration, and scoring of a classroom observation instrument are discussed. The procedures that were used to collect and analyze the observation data are also explained.

Population and Sample

A large urban school division in Virginia with 35 elementary schools, each with a kindergarten through fifth grade configuration was used as the setting for this study. Schools in this division range in size from approximately 200 to over 850 students. Of the 35 elementary schools in the division, 17 are designated as Title I schools because of the high percentage of their students who receive free or reduced price lunch. The schools also receive additional funding and personnel support from state and federal sources. There are approximately 120 classrooms per grade level. For the purpose of this study, the entire population included all of the kindergarten, first grade, and second grade classrooms in the district.

A stratified random sample of this population was used to select the schools that were involved in this study. The purpose of this sampling technique is to have this study sample represent the total population of Norfolk's kindergarten through second grade classrooms. Although the population of this study was generally homogenous, the schools and classrooms selected for this study needed to be proportionally allocated to represent each of the geographic quadrants of the city as well as of the proportion of schools that receive a Title I funding allocation in the school division. Four schools per geographic quadrant were selected, with two of the four schools in each quadrant being Title I schools. The total sample size was 16 schools,

with eight of the schools designated as Title I schools. There were approximately 60 classrooms at each grade level (kindergarten, first and second grades) used as a part of this study.

Instrumentation and Data Collection Procedures

In this section of the chapter, the research questions, observation instrument domains, and validation procedures are explained. The procedures for conducting and scoring the observation instrument are presented as well.

Instrumentation

The research questions for this study are: (1) How can the implementation of a school or district reading program be assessed in a valid, reliable, and efficient way? (2) What accounts for the greatest variance in the implementation of a district's primary literacy program?

This study seeks to answer these questions through a program implementation assessment designed to describe the progress being made in implementing a district wide reading program. The purpose of this study is to develop and test a classroom observation inventory as a formative program implementation tool. This formative study of a district's primary (kindergarten through second grade) literacy program implementation was conducted in an effort to provide information that can be used to improve and strengthen the program and serve as a tool for other school districts that are implementing similar programs.

Development. A classroom observation checklist was developed using the major components of the research-based Norfolk Public Schools Elementary Reading Plan. The checklist items were drawn directly from the plan itself as well as from training sessions held on the implementation of the plan. The checklist is divided into four major sections or domains: (1) learning environment, (2) reading instruction – daily requirements, (3) use of resources, and (4) planning and preparation. The final version of the observation checklist contained 26 items.

The observer(s) looked for each item on the checklist during a brief classroom visit of approximately 15 minutes and assigned it a rating of clearly evident, somewhat evident, or not evident. The observers looked at classroom instruction, classroom environment, lesson plans, and literacy assessment records during their visit to each classroom.

Validity. Content validation of the classroom observation instrument was accomplished as follows. An instrument that contained each of the classroom observation checklist items and response sections for identification of domain and item-domain association strength was constructed (Appendix E). Each of the five domains was defined on the validation instrument. During July and August of 2004, the validation instrument was distributed to a subgroup of doctoral candidates completing their degree requirements in Educational Leadership and Policy Studies at Virginia Polytechnic Institute and State University in Blacksburg as well as via email to a small subgroup of elementary principals employed with Norfolk Public Schools. These schools were not included in this study. They classified each of the statements on the observation checklist as belonging to one of the four domains and also rated each of the statements for its level of association. As a result of this input (Appendix F), all items had a content validity of greater than 80 per cent. No items were eliminated at this point in the study.

In addition, the researcher and a trained partner piloted the classroom observation instrument in the primary classrooms of one Norfolk elementary school. The results of the pilot were used to improve the content and face validity of the instrument. This instrument (Appendix G) was then re-piloted in different primary classrooms in the same school before the instrument was finalized and used for data collection.

Reliability. According to Linn and Gronlund (1998), interrater reliability " is the best method of estimating reliability when attempting to achieve consistency of judgmental scores"

(p. 119). In an effort to ensure consistency of scoring with this particular instrument, each pair of observers visited the first classroom in each school together using a same color-coded observation sheet. Each observer independently scored each of the items on the observation instrument. The observers then correlated their scores with each other in order to achieve "consensus among the observers regarding the types of performances that are valued" with each of the observation checklist items (Linn & Gronlund, p. 117). During training in the use of the instrument, observers were instructed to continue their discussion of their paired ratings until they had achieved at least 85% of the items (22 items out of the 26 on the checklist) in exact agreement. The percentage of exact agreement on the initial observations conducted by the pairs of raters was also calculated (Appendix H). Both of the color-coded observation sheets were turned in with the rest of the observation data, but only one was used in the final tabulation of this study's data. If the scores by the two raters ultimately differed on a given observation checklist item, the average score was used.

Scoring. Each item on the classroom observation checklist was given a rating based on what the observer was able to see or find documentation of during a classroom visit. For each checklist item, the ratings were: clear evidence (2), some evidence (1), and no evidence (0). Quantitative values were assigned to each of these three ratings as indicated by the number in parentheses. A higher score indicates a greater degree of implementation of each observation checklist item. In addition to these ratings, the grade level was noted with a corresponding quantitative value as: kindergarten (0), first grade (1), and second grade (2). Likewise, each classroom was coded as either Title I (1), or non-Title I (0) respectively.

Administration of Instrument

Two person teams of observers were assembled. Each of the team members had some experience in elementary reading instruction (e.g. former primary classroom teacher, former reading specialist, advanced degree in reading, experience as elementary building administrator) and all came from the central office. Team members had no direct line authority over any of the classrooms or buildings that they observed.

All team members attended a two-hour training session in scoring the classroom observation instrument using a structured training agenda (Appendix I). The focus of the training was to obtain a high degree of inter-rater consistency with the classroom observation instrument. By using a well-defined set of guidelines for scoring the instrument and training raters in a systematic way, interrater reliability can be dramatically increased (Linn & Gronlund, 1998, p. 129). During the training, each of the domains and their subsequent instrument items were defined and explained. The explanation included examples of classroom artifacts and teaching that would provide evidence for rating each of the instrument items. Sample resources that classroom observers should see were shared. A commercial video of a primary reading classroom and reading lesson was shown during the training. The video was rated using the observation checklist by each of the participants. Each participant then discussed his individual rating of each item on the checklist with his or her observation partner. Ratings on each of the items were also discussed with the whole group. Other various potential observation scenarios were raised and discussed. A sample script for the teams to use as they entered each building and met with the principal was shared (Appendix J). Team members spent some time reviewing and personalizing the script. Team members also were instructed to observe their first classroom in

each school together to validate their ratings, and then to observe the remaining classes alone. Classrooms with substitute teachers were to be excluded from the observation pool.

The primary classroom observations took place over a two-week period of time. The observation instrument was designed to measure the degree to which each primary classroom is implementing the reading plan components. The observation checklist took approximately 15 minutes per classroom to complete. Each pair of observers was provided with observation sheets (including color-coded sheets to conduct their initial joint observation) as well as a list of classrooms and the time of day that reading instruction was scheduled in each classroom. This short window of observation time reduced some threats to internal validity that could occur primarily through instrument use history – an effort to ensure that the environment in which the study happens in all schools that were selected was the same. Another potential threat in this study was maturation – once administration of the instrument began, the instrument was no longer measuring a neutral environment. In this case, instead of the observation instrument measuring the actual essence of the program, it could instead have been measuring a teacher or a principal's ability to adjust to a school visit that could be forthcoming.

Data Management and Analysis

As the classroom observation sheets were completed and returned, the data was entered into a standard statistical software program (Statistical Package for the Social Sciences, known as SPSS version 11.0) for statistical analyses. Each of the observation checklist variables was entered as well as the grade level of the classroom and the type of school (Title I or non-Title I). One of the same color-coded sheets from the initial observation done in each school was entered. In the few cases where the ratings on individual items differed, an average score of the two ratings was used.

These analyses included the computation of descriptive statistics including means and standard deviations for each item on the checklist for the entire sample. A cross tabs analysis was used to disaggregate the data in several ways. The cross tabs procedure classifies the values of each variable in rows by the values of another variable in columns. For example, the data are reported by observation checklist item and corresponding rating (2, 1, or 0) by grade level. The data were also sorted by observation checklist item rating and designation as a Title I or non-Title I school. Cross tabs analysis was used to report percentages by grade and school designation according to each of the five observation instrument domains.

Chapter Summary

The process for developing, administering and scoring the classroom observation instrument that is the focus of this study was explained in this chapter. In addition, the process that was used to enter and initially analyze the data gathered from the observation instrument was described. The approval to conduct this particular study was granted by the local Virginia school division in which this study was conducted through their department of research, testing and statistics (Appendix K) upon submission of the abstract and chapter three of this study. The Institutional Review Board of the Virginia Polytechnic Institute and State University agreed that the research described in this study fell within the exempt status guidelines and granted IRB exempt approval for the study (Appendix L).

CHAPTER IV

RESULTS

The purpose of this study was to develop and test a classroom observation inventory as a formative reading program implementation tool. This chapter reports the results of the classroom observations conducted by the observation teams using the classroom observation inventory. A review of the procedures followed by the observers as well as the actual sample population precedes the observation inventory results.

Observation Checklist Implementation

A classroom observation checklist was developed using the major components of a research-based school division elementary reading plan as a means of answering research question one (How can the implementation of a school or district reading program be assessed in a valid, reliable, and efficient way?). The checklist items were drawn directly from the plan itself as well as from training sessions the division held on the implementation of the plan. The checklist was divided into four sections or domains: (1) learning environment, (2) reading instruction – daily requirements, (3) use of resources, and (4) planning and preparation for reading instruction. The initial instrument was tested for content and face validity using the procedures described in chapter three of this study. The final version of the checklist contained 26 items.

Sixteen observers were trained in scoring the classroom observation instrument using a structured training agenda (Appendix I) in a two-hour session. The observers were selected from the central office staff because they had some experience with elementary reading instruction (e.g. former reading specialist, advanced degree in reading, former primary classroom teacher,

former elementary principal). None of the team members had direct line authority over any of the classrooms or buildings that they observed.

The focus of the training was to obtain a high degree of inter-rater consistency with the instrument. During the training, each of the domains and their subsequent instrument items were defined and explained using examples of classroom artifacts and teaching that provided evidence for rating each of the instrument items. Sample resources that the observers should see were also shared. A commercial video of a primary reading classroom and reading lesson was shown. Each of the participants rated the lesson using the instrument checklist. The team also reviewed and personalized an introductory script (Appendix J) for use as they entered buildings and met with principals.

Following the training session described in chapter three, schools that were selected using random stratified sampling were alphabetically assigned to pairs of observers. The observers were given a two-week window to conduct their observations using the classroom observation checklist. During the designated two-week window, the classroom observation checklist was used by eight pairs of observers in 182 primary classrooms in fifteen elementary schools. Of the 182 classrooms that were ultimately observed, 66 were kindergarten classrooms, 60 were first grade classrooms, and 56 were second grade classrooms. While the original study design called for the observations to take place in 16 elementary schools representing approximately 180 classrooms (60 at each grade level), 15 elementary schools representing approximately the same estimated sample size were ultimately used. The classroom observations in the sixteenth school could not be conducted by the assigned observers within the designated observation window. Without the sixteenth school, the total number of classrooms observed (182) exceeded the number estimated in the study design that would be observed (180). In all, the total number of

Title One classrooms visited was 92, and the total number of non Title One classes visited was 90 (Table 3) with the number of classrooms within each school type evenly distributed.

Table 3

Sample Size by Grade Level and School Type

	Grade Level			
	Kindergarten	First Grade	Second Grade	Total
Title One	32	31	29	92
Non Title One	34	29	27	90
Total	66	60	56	182

Results of the Implementation of the Observation Instrument

Each pair of observers was provided with observation sheets as well as with a list of the classrooms and the time of day that reading instruction was scheduled for each grade level. Observers used a color-coded sheet to conduct their initial observation at each school as a joint observation, and then observed the other primary classrooms in each of their two schools individually. Classrooms with substitute teachers present on the day of the visit were not observed.

For each of the 182 classrooms that were eventually observed using the checklist, the grade level of the classroom and the type of school (Title I or non-Title I) was coded. A rating of either “2” (clearly evident), “1” (somewhat evident), or “0” (not evident) was then assigned to each of the 26 items on the instrument. The frequency and percent of the entire sample of observed classrooms that received each of the ratings (Table 4) was then calculated and organized by each of the four instrument sect5ions or domains.

Table 4

Frequency and Percent by Item of Entire Sample

Item Number	Frequency/Percent by Rating		
	2	1	0
Domain: Environment/ Print			
1 – Work Area	136/ 75%	42/23%	4/2%
2 – Routines	80/44%	66/36%	36/20%
3 – Organization	112/62%	66/36%	4/2%
4 – Print Rich	85/47%	80/44%	17/9%
5 – Word Wall	93/51%	48/26%	41/23%
6 – Student Work	31/17%	53/29%	98/54%
7 – Classroom Library	81/44%	72/40%	29/16%
8 – BTL in Use	53/82%	8/12%	4/6%
Domain: Literacy/ Reading Instruction			
9 – Two Hours	133/73%	2/1%	47/26%
10 – Shared Reading	111/61%	35/19%	36/20%
11 – Guided Reading	95/52%	41/23%	46/25%
12 – Word Study	85/47%	57/31%	40/22%
13 – Whole Group Focus	93/51%	51/28%	38/21%
14 – Small Group Focus	76/42%	55/30%	51/28%
15 – Centers	78/43%	32/18%	72/40%
16 – Center Work	55/30%	57/31%	70/39%
17 – No Worksheets	68/37%	55/30%	59/32%

18 – Portfolios	122/67%	34/19%	26/14%
Domain: Resources			
19 – BTL Books	48/74%	13/20%	4/6%
20 – Basal	124/68%	19/10%	39/21%
21 – Leveled Books	100/51%	41/22.5%	41/22.5%
22 – Big Books	103/57%	21/12%	58/32%
Domain: Planning/ Preparation			
23 – Lesson Plans	142/78%	28/15%	12/7%
24 – Plans for All	75/41%	84/46%	23/13%
25 – Adults w/Students	144/79%	21/12%	17/9%
26 – Reading Level	77/42%	41/23%	64/35%

Note. Ratings for each checklist items were scored on a 3 point scale (2 = clearly evident, 1 = some evidence, and 0 = no evidence).

The checklist items were "clearly evident" in at least 40% of the classrooms observed. With a few exceptions, there was "no evidence" of checklist items in less than 30% of the classrooms that were observed. Generally, checklist items that received a rating of "somewhat evident" were in the 10 to 25 percent range.

Each observer assigned a rating of either "2" (clearly evident), "1" (somewhat evident), or "0" (not evident) to each of the 26 items on the checklist. The mean score for each of the 26 items on the K –2 observation checklist ranged from a low of .6319 for checklist item number six (students' original, current work supporting reading and writing is posted) to a high of 1.7538 for checklist item number eight (Breakthrough to Literacy program is in use in kindergarten classrooms). The checklist item with the second highest score (item one, well-defined areas for

whole group instruction and small group work) had the lowest standard deviation (.49452); however, the item with the greatest standard deviation was checklist item 22 (use of big books as a teaching resource). Table five displays the mean score and standard deviation for each of the checklist items.

Table 5

Mean Score and Standard Deviation for Checklist Items

Item	Mean (Range 0 –2)	Standard Deviation
Domain: Environment/ Print		
1 – Work Area	1.7253	.49452
2 – Routines	1.2418	.76296
3 – Organization	1.5934	.53554
4 – Print Rich	1.3736	.65051
5 – Word Wall	1.2857	.81132
6 – Student Work	.6319	.75923
7 – Classroom Library	1.2857	.72502
8 – BTL in Use	1.7538	.55988
Domain: Literacy/Reading Instruction		
9 – Two Hours	1.4725	.87747
10 – Shared Reading	1.4121	.80087
11 – Guided Reading	1.2692	.84031
12 – Word Study	1.2473	.79318
13 – Whole Group Focus	1.3022	.79494

14 – Small Group Focus	1.1374	.82625
15 – Centers	1.0330	.90975
16 – Center Work	.9176	.82691
17 – No Worksheets	1.0495	.83618
18 – Portfolios	1.5275	.73343
Domain: Resources		
19 – BTL Books	1.6769	.58916
20 – Basal	1.4670	.82537
21 – Leveled Books	1.3242	.82057
22 – Big Books	1.2473	.90996
Domain: Planning/ Preparation		
23 – Lesson Plans	1.7143	.58121
24 – Plans for All	1.2857	.67776
25 – Adults w/Students	1.6978	.63236
26 – Reading Level	1.0714	.87970

The low number of possible checklist ratings (three) contributed to many of the high and low standard deviations. For example, most of the items in the "resources" domain had high standard deviations. In these cases, either the teachers had and were using the resources, or they did not have them and were not using them. Print resources cannot be "somewhat evident". The same is true of checklist item 9 (at least two hours of reading instruction). Either the teachers in the observed classrooms had the scheduled two hours, or they did not. In measuring scheduled time, it is virtually impossible to receive a rating of "1" (somewhat evident). Similar results

occurred with checklist items 15 and 16 (use of centers and type of work in the centers). With both of these instrument items, either teachers had centers and the subsequent appropriate center work, or they did not. For example, it would be impossible to have "appropriate center work" in a classroom that did not have centers. On the other hand, in the case of the use of Breakthrough to Literacy print resources (checklist item 19), the low standard deviation paralleled the high mean score. Most of the kindergarten teachers in the classrooms observed using the checklist received a rating of "2" (clearly evident) on this particular checklist item.

Results by Classification of School

In addition to providing a rating for each of the instrument items, each observer also indicated on the observation instrument whether the classroom was in a Title One or a non-Title One school by coding either a "0" (for non-Title One) or a "1" (for Title One) at the top of the instrument. The frequency and percent of instrument items that received each of the ratings in all Title One classrooms was then calculated (Table 6) as one way of answering research question two (What accounts for the greatest degree of variance in the implementation of a district's primary literacy program?). These frequencies are organized by each of the four domains that were on the classroom observation instrument.

Table 6

Frequency and Percent by Item in Title I Schools

Item Number	Frequency/Percent by Rating		
	2	1	0
Domain: Environment/ Print			
1 – Work Area	72/78%	18/20%	2/2%
2 – Routines	40/44%	27/29%	25/27%

3 – Organization	61/66%	30/33%	1/1%
4 – Print Rich	33/36%	47/51%	12/13%
5 – Word Wall	53/58%	27/29%	12/13%
6 – Student Work	12/13%	26/28%	54/59%
7 – Classroom Library	46/50%	28/30%	18/20%
8 – BTL in Use	26/81%	3/9%	3/9%
Domain: Literacy/ Reading Instruction			
9 – Two Hours	64/70%	1/1%	27/29%
10 – Shared Reading	62/67%	14/15%	16/17%
11 – Guided Reading	52/57%	20/22%	20/22%
12 – Word Study	58/63%	17/18%	17/18%
13 – Whole Group Focus	49/53%	23/25%	20/22%
14 – Small Group Focus	35/38%	30/33%	27/29%
15 – Centers	45/49%	16/17%	31/34%
16 – Center Work	30/33%	34/37%	28/30%
17 – No Worksheets	37/40%	29/31%	26/28%
18 – Portfolios	61/66%	18/20%	13/14%
Domain: Resources			
19 – BTL Books	24/74%	4/13%	4/13%
20 – Basal	61/66%	6/6%	25/27%
21 – Leveled Books	57/62%	20/22%	15/16%
22 – Big Books	52/57%	9/10%	31/34%
Domain: Planning/ Preparation			

23 – Lesson Plans	75/82%	14/15%	3/3%
24 – Plans for All	45/49%	34/37%	13/14%
25 – Adults w/Students	72/78%	10/11%	10/11%
26 – Reading Level	49/53%	22/24%	21/23%

Note. Ratings for each checklist items were scored on a 3 point scale (2 = clearly evident, 1 = some evidence, and 0 = no evidence).

The highest levels of clear evidence in Title One classrooms included item 23 (teachers had lesson plans for reading), item 8 (use of the Breakthrough to Literacy program in kindergarten classrooms), item 1 (clearly defined areas for whole and small group work), and item 25 (all adults in the classroom were clearly working with students on reading instruction). The lowest levels of implementation were in item 6 (displays of student work), item 34 (use of big books as a resource), and item 15 (use of centers for independent work).

The frequency and percent of instrument items that received each of the ratings in all non-Title One classrooms was also calculated (Table 7). These frequencies are organized by each of the four domains that were on the classroom observation instrument as well.

Table 7

Frequency and Percent by Item in Non-Title I Schools

Item Number	Frequency/Percent by Rating		
	2	1	0
Domain: Environment/ Print			
1 – Work Area	64/71%	24/27%	2/2%
2 – Routines	40/44%	39/43%	11/12%
3 – Organization	51/57%	36/39%	3/3%

4 – Print Rich	52/58%	33/37%	5/6%
5 – Word Wall	40/44%	21/23%	29/32%
6 – Student Work	19/21%	27/30%	44/49%
7 – Classroom Library	35/39%	44/48%	11/12%
8 – BTL in Use	27/82%	5/15%	1/3%
Domain: Literacy/ Reading Instruction			
9 – Two Hours	69/77%	1/1%	20/22%
10 – Shared Reading	49/54%	21/23%	20/22%
11 – Guided Reading	43/48%	21/23%	26/29%
12 – Word Study	27/30%	40/45%	23/25%
13 – Whole Group Focus	44/49%	28/31%	18/20%
14 – Small Group Focus	41/46%	25/28%	24/27%
15 – Centers	33/37%	16/18%	41/46%
16 – Center Work	25/28%	23/26%	42/47%
17 – No Worksheets	31/33%	26/29%	33/37%
18 – Portfolios	61/69%	16/17%	13/15%
Domain: Resources			
19 – BTL Books	24/73%	9/27%	0/0%
20 – Basal	63/70%	13/14%	14/16%
21 – Leveled Books	43/48%	21/23%	26/29%
22 – Big Books	51/57%	12/13%	27/30%
Domain: Planning/ Preparation			
23 – Lesson Plans	67/74%	14/16%	9/10%

24 – Plans for All	30/33%	50/56%	10/11%
25 – Adults w/Students	72/80%	11/12%	7/8%
26 – Reading Level	28/31%	19/21%	43/48%

Note. Ratings for each checklist items were scored on a 3 point scale (2 = clearly evident, 1 = some evidence, and 0 = no evidence).

The highest levels of clear evidence in the non Title One classes included item 8 (use of the Breakthrough to Literacy program in kindergarten classrooms), item 25 (all adults in the classroom were clearly working with students on reading instruction), and item 9 (evidence of at least 2 hours of literacy instruction). The lowest levels of implementation were in item 6 (displays of student work), item 26 (planning that incorporated student reading levels), and item 47 (type of independent student work).

The frequency and percent of instrument items that received each of the ratings in all non-Title One classrooms as well as in the Title One classrooms was also calculated (Table 8). These frequencies are organized by each of the four domains that were on the classroom observation instrument and show a comparison of the level of implementation for each instrument item by school classification.

Table 8

Frequency and Percent by Rating in Non-Title I and Title I Schools

Item Number	Frequency/Percent by Rating					
	2		1		0	
	Ttl I	Non Ttl I	Ttl I	Non Ttl I	Ttl I	Non Ttl I
Domain: Environment/ Print						
1 – Work Area	72 78%	64 71%	18 20%	24 27%	2 2%	2 2%

2 – Routines	40	40	27	39	25	11
	44%	44%	29%	43%	27%	12%
3 – Organization	61	51	30	36	1	3
	66%	57%	33%	39%	1%	3%
4 – Print Rich	33	52	47	33	12	5
	36%	58%	51%	37%	13%	6%
5 – Word Wall	53	40	27	21	12	29
	58%	44%	29%	23%	13%	32%
6 – Student Work	12	19	26	27	54	44
	13%	21%	28%	30%	59%	49%
7 – Classroom Library	46	35	28	44	18	11
	50%	39%	30%	48%	20%	12%
8 – BTL in Use	26	27	3	5	3	1
	81%	82%	9%	15%	9%	3%
Domain: Literacy/ Reading Instruction						
9 – Two Hours	64	69	1	1	27	20
	70%	77%	1%	1%	29%	22%
10 – Shared Reading	62	49	14	21	16	20
	67%	54%	15%	23%	17%	22%
11 – Guided Reading	52	43	20	21	20	26
	57%	48%	22%	23%	22%	29%
12 – Word Study	58	27	17	40	17	23
	63%	30%	18%	45%	18%	25%
13 – Whole Group Focus	49	44	23	28	20	18
	53%	49%	25%	31%	22%	20%
14 – Small Group Focus	35	41	30	25	27	24
	38%	46%	33%	28%	29%	27%
15 – Centers	45	33	16	16	31	41
	49%	37%	17%	18%	34%	46%
16 – Center Work	30	25	34	23	28	42
	33%	28%	37%	26%	30%	47%
17 – No Worksheets	37	31	29	26	26	33
	40%	33%	31%	29%	28%	37%
18 – Portfolios	61	61	18	16	13	13
	66%	69%	20%	17%	14%	15%
Domain: Resources						
19 – BTL Books	24	24	4	9	4	0
	74%	73%	13%	27%	13%	0%
20 – Basal	61	63	6	13	25	14
	66%	70%	6%	14%	27%	16%
21 – Leveled Books	57	43	20	21	15	26
	62%	48%	22%	23%	16%	29%
22 – Big Books	52	51	9	12	31	27
	57%	57%	10%	13%	34%	30%

Domain: Planning/ Preparation						
23 – Lesson Plans	75 82%	67 74%	14 15%	14 16%	3 3%	9 10%
24 – Plans for All	45 49%	30 33%	34 37%	50 56%	13 14%	10 11%
25 – Adults w/Students	72 78%	72 80%	10 11%	11 12%	10 11%	7 8%
26 – Reading Level	49 53%	28 31%	22 24%	19 21%	21 23%	43 48%

Note. Ratings for each checklist items were scored on a 3-point scale (2 = clearly evident, 1 = some evidence, and 0 = no evidence).

Within each of the four domains on the instrument, Title One classrooms had equal or greater percentages of "clearly evident" ratings than non-Title One classrooms on several instrument items. Six of the eight instrument items in the environment/print domain and seven of the ten instrument items in the literacy/ reading instruction domain had percentages of "clearly evident" that were equal to or great than those of the non-Title One classrooms. Similarly, three of the four checklist items in the resources domain, and all four of the checklist items in the planning/ preparation domain had percentages of "clearly evident" in Title One classrooms that were equal to or better than those in the non-Title One classrooms. The greatest difference in percentage of "clearly evident" between the Title One and non-Title One classrooms was checklist item 12 (word study) which was "clearly evident" in 63% of the Title One classes and was "clearly evident" in 30% of the non-Title One classes.

The non-Title One classrooms observed using the instrument received an equal or greater percentage of ratings of no evidence ("0") on nine of the ten checklist items in the literacy/ reading instruction domain. The greatest difference in percentage of "no evidence" ratings was checklist item 26 (use of reading level), which was not evident in 23% of the Title One classrooms while it was not evident in 48% of the non-Title One classrooms.

Results by Frequency and Percent

Using the cross tabs analysis procedure, data from the classroom observation survey was also sorted within each of the checklist ratings to determine which individual checklist items received both the largest number as well as the highest percentage of ratings of "2" (clearly evident) in the entire sample as well as in both the Title One and the non-Title One classrooms. Frequencies for each of the instrument items were initially examined; however, two of the instrument items were applicable only to kindergarten classrooms so percentage scores were ultimately used to determine the highest and lowest scoring instrument items. In reporting the data, a cut score of at least 70 percent on each instrument item was initially used to examine the data, and then the data were narrowed further with a cut score of at least 80 percent on the individual instrument items.

In the entire sample of kindergarten through second grade classrooms observed, six of the checklist items had a rating of "2" (clearly evident) in greater than 70 percent. One checklist item (use of Breakthrough to Literacy program) had a rating of "2" in more than 80 percent of the entire sample of classrooms observed (Table 9).

Table 9

Top Items with a Score of "2" from Entire Sample

Item	Domain	Frequency/ Percent
8 – Use of Breakthrough	Environment/ Print	53/ 82%
25 – Adults Work w/ Students	Planning/ Preparation	144/ 79%
23 – Lesson Plans	Planning/ Preparation	142/ 78%
1 – Work Area	Environment/ Print	136/ 75%
19 – BTL Books	Resources	48/ 74%

9 – Two Hours of Reading	Literacy/Reading Instruction	133/73%

Checklist items from the entire sample that received the greatest number and percentage of scores with a "2" were from each of the four instrument domains. The percent of classrooms where these descriptors were clearly evident ranged from 73 to 82 percent. For example, the instrument item with the highest rating of "clearly evident" from the entire sample was checklist item 8 (use of Breakthrough), which was clearly evident in 53 (or 82 percent) of the kindergarten classrooms that were observed as part of this study.

In the Title One classrooms that were observed using the checklist, five of the items had a rating of "2" (clearly evident) in more than 70 percent of the observations. Two checklist items had a rating of "2" in greater than 80 percent of the Title One classrooms observed (Table 10).

Table 10

Top Items with a Score of "2" in Title I Schools

Item	Domain	Frequency/ Percent
23 – Lesson Plans	Planning/ Preparation	75/ 82%
8 – Use of Breakthrough	Environment/ Print	26/ 81%
25 – Adults Work w/ Students	Planning/ Preparation	72/ 78%
1 – Work Area	Environment/ Print	72/ 78%
19 – BTL Books	Resources	24/ 74%

Checklist items that received the greatest number and percentage of scores with a "2" were from three of the four instrument domains. The percent of classrooms where these descriptors were clearly evident ranged from 74 to 82 percent. The top items with a score of "2"

were from three of the four domains on the classroom observation instrument. The instrument item with the top score of clearly evident in the Title One classrooms was the same as that from the entire sample (instrument item 8, use of Breakthrough to Literacy). The domain not represented with instrument item scores at or above 70 percent was literacy/reading instruction.

In the non-Title One classrooms that were observed using the checklist, six of the checklist items had a rating of "2" (clearly evident) in more than 70 percent of the kindergarten through second grade classrooms observed. Two checklist items had a rating of "2" in more than 80 percent of the entire sample of classrooms observed (Table 11).

Table 11

Top Items with a Score of "2" in Non-Title I Schools

Item	Domain	Frequency/ Percent
8 – Use of Breakthrough	Environment/ Print	27/ 82%
25 – Adults Work w/ Students	Planning/ Preparation	72/ 80%
9 – Two Hours of Reading	Literacy/Reading Instruction	69/ 77%
23 – Lesson Plans	Planning/ Preparation	67/ 74%
19 – BTL Books	Resources	24/ 73%
1 – Work Area	Environment/ Print	64/ 71%

Checklist items that received the greatest number and percentage of scores with a "2" were from each of the four instrument domains. The percent of classrooms where these descriptors were clearly evident ranged from 71 to 82 percent. The checklist items that received the highest score in the non-Title One classrooms observed was the same as that in the entire

sample as well as in the Title One classrooms (instrument item 8, use of the Breakthrough to Literacy program) which was rated clearly evident in 27 (82%) of the non-Title One classrooms.

Using the cross tabs analysis procedure, data from the classroom observation survey was also sorted within each of the checklist ratings to determine which individual checklist items received both the largest number as well as the highest percentage of ratings of "0" (no evidence) in the entire sample as well as in both the Title One and the non-Title One classrooms. In reporting the data, a cut score of at least 30 percent of the classrooms showing no evidence on each instrument item was initially used to examine the data, and then the data was narrowed further with a cut score of at least 50 percent of the classrooms showing no evidence on the individual instrument items.

In the entire sample of kindergarten through second grade classrooms, six of the checklist items had a rating of "0" (no evidence) in more than 30 percent of the observed classrooms. One of these checklist items, students' original work supporting reading and writing is posted, had a rating of "0" in more than 50 percent of the entire sample of classrooms observed (Table 12).

Table 12

Top Items with a Score of "0" from Entire Sample

Item	Domain	Frequency/ Percent
6 – Student Work	Environment/ Print	98/ 54%
15 – Centers	Literacy/ Reading Instruction	72/ 40%
16 – Center Work	Literacy/ Reading Instruction	70/ 39%
26 – Groups by Reading Level	Planning/ Preparation	64/ 35%
17 – No Worksheets	Literacy/ Reading Instruction	59/ 32%
22 – Use of Big Books	Resources	58/ 32%

Checklist items that received the greatest number and percentage of scores with a "0" were from each of the four instrument domains. The percent of classrooms where there was no evidence of these descriptors ranged from 32 to 54 percent.

In the Title One classrooms observed, three of the checklist items had a rating of "0" (no evidence) in more than 30 per cent of the classrooms. One of these checklist items had a rating of "0" in more than 50 percent of the entire sample of classrooms observed (Table 13).

Table 13

Top Items with a Score of "0" in Title One Schools

Item	Domain	Frequency/ Percent
6 – Student Work	Environment/ Print	54/ 59%
15 – Centers	Literacy/ Reading Instruction	31/ 34%
22 – Use of Big Books	Resources	31/ 34%

Checklist items that received the greatest number and percentage of scores with a "0" in the Title One classrooms observed were from three of the four instrument domains. The domain not represented was planning/ preparation. The percent of classrooms where there was no evidence of these descriptors ranged from 34 percent for two of the instrument items to 59 percent for the third. The items with the highest percent of no evidence in the Title One classrooms was the same as that of the entire population.

In the non-Title One classrooms observed, six of the checklist items had a rating of "0" (no evidence) in more than 30 percent of the classrooms. None of the checklist items had a rating of "0" in more than 50 percent of the entire sample of classrooms observed (Table 14).

Table 14

Top Five Items with a Score of "0" from Non-Title One Schools

Item	Domain	Frequency/ Percent
6 – Student Work	Environment/ Print	44/ 49%
26 – Groups by Reading Level	Planning/ Preparation	43/ 48%
16 – Center Work	Literacy/ Reading Instruction	42/ 47%
15 – Centers	Literacy/ Reading Instruction	41/ 46%
17 – No Worksheets	Literacy/ Reading Instruction	33/ 37%

Checklist items that received the greatest number and percentage of scores with a "0" (Table 14) were from three of the four instrument domains. The domain not represented was resources, while the domain not represented in the Title One classrooms was planning/ preparation. The percentage of classrooms where there was no evidence of these descriptors ranged from 37 to 49 percent. Checklist item 6, students' original work supporting reading and writing is posted, had the highest percent of no evidence in the non-Title One classrooms which mirrored the results of the entire sample as well as the sample of Title One classrooms.

Results by Grade Level Designation

In addition to looking at the classroom observation and checklist data according to school type (Title One or non-Title One) to determine the degree of implementation of a district's primary literacy program, data were also gathered by primary grade level designation (kindergarten, first grade or second grade).

A total of 66 kindergarten classrooms were observed as part of this study. Of the 66 classrooms, 32 were Title One kindergarten classes, and 34 were non-Title One kindergarten

classes. The frequency and percent for each of the items on the observation checklist in these kindergarten classrooms was determined and organized by instrument domain (Table 15).

Table 15

Frequency and Percent by Item in Kindergarten Classrooms

Item Number	Frequency/Percent by Rating		
	2	1	0
Domain: Environment/ Print			
1 – Work Area	54/32%	11/17%	1/1%
2 – Routines	23/35%	26/39%	17/26%
3 – Organization	44/68%	21/31%	1/1%
4 – Print Rich	30/46%	27/41%	9/14%
5 – Word Wall	27/41%	19/29%	20/30%
6 – Student Work	10/15%	24/36%	32/49%
7 – Classroom Library	28/43%	26/40%	11/17%
8 – BTL in Use	53/82%	8/12%	4/6%
Domain: Literacy/ Reading Instruction			
9 – Two Hours	45/68%	0	21/32%
10 – Shared Reading	42/64%	18/27%	6/9%
11 – Guided Reading	26/39%	16/24%	24/36%
12 – Word Study	25/39%	26/40%	15/21%
13 – Whole Group Focus	30/46%	25/38%	11/17%
14 – Small Group Focus	21/32%	21/32%	24/36%
15 – Centers	24/36%	13/20%	29/44%

16 – Center Work	20/30%	19/29%	27/41%
17 – No Worksheets	26/38%	24/36%	16/24%
18 – Portfolios	41/63%	15/23%	9/14%
Domain: Resources			
19 – BTL Books	48/74%	13/20%	4/6%
20 – Basal	34/52%	10/15%	22/33%
21 – Leveled Books	42/64%	14/21%	10/15%
22 – Big Books	60/91%	3 /4%	3 /4%
Domain: Planning/ Preparation			
23 – Lesson Plans	47/71%	12/18%	7/11%
24 – Plans for All	31/47%	24/36%	11/17%
25 – Adults w/Students	49/74%	11/17%	6/9%
26 – Reading Level	20/30%	15/23%	31/47%

Note. Ratings for each checklist items were scored on a 3 point scale (2 = clearly evident, 1 = some evidence, and 0 = no evidence).

Of the 26 items on the instrument checklist, item 22 (use of big books) had the greatest degree of clear evidence. Checklist item 6 (students' original, current work supporting reading/writing is posted) had the greatest degree of no evidence.

There were 60 first grade classrooms that were observed as part of this study. Of this number, 31 were in Title One schools, and 29 were in non-Title One schools. The results of all checklist items for the first grade classrooms that were observed as part of this study are arranged by domain (Table 16) and by item rating within each of the four domains.

Table 16

Frequency and Percent by Item in First Grade Classrooms

Item Number	Frequency/Percent by Rating		
	2	1	0
Domain: Environment/ Print			
1 – Work Area	45/75%	14/23%	1/ 2%
2 – Routines	31/52%	16/27%	13/22%
3 – Organization	42/70%	17/28%	1/ 2%
4 – Print Rich	30/50%	28/47%	2 /3%
5 – Word Wall	37/62%	17/28%	6/10%
6 – Student Work	12/20%	14/23%	34/57%
7 – Classroom Library	28/43%	26/40%	11/17%
8 – BTL in Use	0	0	0
Domain: Literacy/ Reading Instruction			
9 – Two Hours	47/78%	2 / 3%	11/18%
10 – Shared Reading	42/70%	8/13%	10/17%
11 – Guided Reading	36/60%	13/22%	11/18%
12 – Word Study	35/58%	14/23%	11/18%
13 – Whole Group Focus	32/53%	13/22%	15/25%
14 – Small Group Focus	31/52%	16/27%	13/22%
15 – Centers	31/52%	11/18%	18/30%
16 – Center Work	18/30%	20/33%	22/37%
17 – No Worksheets	22/37%	12/20%	26/43%

18 – Portfolios	36/60%	12/20%	12/20%
Domain: Resources			
19 – BTL Books	0	0	0
20 – Basal	45/75%	6/10%	9/15%
21 – Leveled Books	34/57%	12/20%	14/23%
22 – Big Books	32/53%	13/22%	15/25%
Domain: Planning/ Preparation			
23 – Lesson Plans	48/80%	9/15%	3/5%
24 – Plans for All	26/43%	31/52%	3/5%
25 – Adults w/Students	50/83%	7/12%	3/5%
26 – Reading Level	30/50%	15/25%	15/25%

Note. Ratings for each checklist items were scored on a 3 point scale (2 = clearly evident, 1 = some evidence, and 0 = no evidence).

The observation checklist item with the highest number and percentage of clear evidence in the first grade classrooms observed was checklist item 25 (all teachers assigned to the classroom work directly with students during all components of reading instruction). The item with the highest number and percentage of no evidence was checklist item number 6 (students' original, current work supporting reading/writing is posted).

There were 56 second grade classrooms that also were observed as part of this study. Approximately one half of the classrooms (29) were in Title One schools, and 27 were in non-Title One schools. All of the checklist items for the second grade classrooms were organized by domain (Table 17) and rating within each of the four domains on the checklist.

Table 17

Frequency and Percent by Item in Second Grade Classrooms

Item Number	Frequency/Percent by Rating		
	2	1	0
Domain: Environment/ Print			
1 – Work Area	37/66%	17/30%	2/ 4%
2 – Routines	26/46%	24/43%	6/11%
3 – Organization	26/46%	28/50%	2/ 4%
4 – Print Rich	25/45%	25/45%	6/10%
5 – Word Wall	29/52%	12/21%	15/27%
6 – Student Work	9/16%	15/27%	32/57%
7 – Classroom Library	25/45%	24/43%	7/12%
8 – BTL in Use	0	0	0
Domain: Literacy/ Reading Instruction			
9 – Two Hours	41/73%	0	15/27%
10 – Shared Reading	27/48%	9/16%	20/36%
11 – Guided Reading	33/59%	12/21%	11/20%
12 – Word Study	25/45%	17/30%	14/25%
13 – Whole Group Focus	31/55%	13/23%	12/21%
14 – Small Group Focus	24/43%	18/32%	14/25%
15 – Centers	23/41%	8/14%	25/45%
16 – Center Work	17/30%	18/32%	21/38%
17 – No Worksheets	20/36%	19/34%	17/30%

18 – Portfolios	45/80%	6/11%	5/9%
Domain: Resources			
19 – BTL Books	0	0	0
20 – Basal	45/80%	3/ 5%	8/14%
21 – Leveled Books	24/43%	15/27%	17/30%
22 – Big Books	11/20%	5/9%	40/71%
Domain: Planning/ Preparation			
23 – Lesson Plans	47/84%	7/12%	2/ 4%
24 – Plans for All	18/32%	29/52%	9/16%
25 – Adults w/Students	45/80%	3/ 5%	8/14%
26 – Reading Level	27/48%	11/20%	18/32%

Note. Ratings for each checklist items were scored on a 3 point scale (2 = clearly evident, 1 = some evidence, and 0 = no evidence).

In the second grade classrooms that were observed using the checklist, item 23 (all teachers assigned to the classroom have individual lesson plans for the day) had the highest number of ratings of "2" (clearly evident). Checklist item number 22 (use of big books as a teaching resource) had the highest number of ratings of "0" (no evidence) in the second grade classrooms. This differed from the kindergarten and first grade samples where in both cases checklist item 6 (students original work supporting reading and writing is posted) had the highest percentages of no evidence.

Checklist items that received the greatest number and percentage of scores with a "2" in the kindergarten classrooms that were observed (Table 18) were from three of the four instrument domains. In reporting this data, a cut score of at least 70 percent on each instrument

item was initially used to examine the grade level data, and then the data was narrowed further with a cut score of at least 80 percent on the individual instrument items.

Table 18

Top Items with a Score of "2" in Kindergarten Classes

Item	Domain	Frequency/ Percent
22 – Big Books	Resources	60/ 91%
1 – Work Area	Environment/ Print	54/ 82%
8 – Use of Breakthrough	Environment/ Print	53/ 82%
19 – BTL Books	Resources	48/ 74%
25 – Teachers with Students	Planning/ Preparation	49/ 74%

N = 66 Kindergarten Classrooms

In the kindergarten classrooms that were observed using the checklist, five of the checklist items had a rating of "2" (clearly evident) in more than 70 per cent of the classrooms observed. Two of these checklist items had a rating of "2" in more than 80 percent of the entire sample of kindergarten classrooms. The domain not included was literacy/reading instruction. The percent of kindergarten classrooms where these descriptors were clearly evident ranged from two checklist items with 74 percent to 91 percent.

In the first grade classrooms that were observed using the checklist, five of the checklist items also had a rating of "2" (clearly evident) in more than 70 per cent of the classrooms observed. Two of these checklist items had a rating of "2" in more than 80 percent of the entire sample of first grade classrooms (Table 19).

Like the kindergarten data, the checklist items from the first grade classes that received the greatest number and percentage of scores with a "2" were from three of the four instrument

Table 19

Top Items with a Score of "2" in First Grade Classes

Item	Domain	Frequency/ Percent
25 – Teachers with Students	Planning/ Preparation	50/ 83%
23 – Teachers Have Plans	Planning/ Preparation	48/ 80%
9 – Two Hours of Reading	Literacy/ Reading Instruction	47/ 78%
1 – Work Area	Environment/ Print	45/ 75%
20 – Current Basal	Resources	45/ 75%

N = 60 First Grade Classrooms

domains. The domain not represented is literacy/reading instruction. The percent of first grade classrooms where these descriptors were clearly evident ranged from 75 to 83 percent.

In the second grade classrooms that were observed using the checklist, five of the checklist items also had a rating of "2" (clearly evident) in more than 70 per cent of the classrooms observed. One of these checklist items had a rating of "2" in more than 80 percent of the entire sample of second grade classrooms (Table 20).

Table 20

Top Items with a Score of "2" in Second Grade Classes

Item	Domain	Frequency/ Percent
23 – Teachers Have Plans	Planning/ Preparation	47/ 84%
18 – Portfolio Assessment	Literacy/ Reading Instruction	45/ 80%
20 – Current Basal	Resources	45/ 80%
25 – Teachers with Students	Planning/ Preparation	45/ 80%
9 – Two Hours of Reading	Literacy/ Reading Instruction	41/ 73%

N = 56 Second Grade Classrooms

Checklist items that received the greatest number and percentage of scores with a "2" were from three of the four instrument domains. The domain not included is environment/ print. The percent of second grade classrooms where these descriptors were clearly evident ranged from 73 to 84 percent.

Using the cross tabs analysis procedure, data from the classroom observation survey was also sorted within each of the checklist ratings to determine which individual checklist items received both the largest number as well as the highest percentage of ratings of "0" (no evidence) in the kindergarten, first grade, and second grade classrooms. In reporting the data, a cut score of at least 30 percent of the classrooms showing no evidence on each instrument item was initially used to examine the data, and then the data were narrowed further with a cut score of at least 50 percent of the classrooms showing no evidence on the individual instrument items.

In the kindergarten classrooms observed, eight of the checklist items had a rating of "0" (no evidence) in greater that 30 per cent of the classrooms; however none of the checklist items had a rating of "0" in greater than 50 percent of the entire sample of kindergarten classrooms observed (Table 21).

Checklist items that received the greatest number and percentage of scores with a "0" in the kindergarten classrooms were from all four of the instrument domains; however, five of the eight instrument items were from the literacy/ reading instruction domain. The percentage of

Table 21

Top Items with a Score of "0" in Kindergarten Classes

Item	Domain	Frequency/ Percent
6 – Student Work	Environment/ Print	32/ 49%

26 – Groups by Reading Level	Planning/ Preparation	31/ 47%
15 – Centers	Literacy/ Reading Instruction	29/ 44%
16 – Center Work	Literacy/ Reading Instruction	27/ 41%
11 – Guided Reading	Literacy/ Reading Instruction	24/ 36%
14 – Small Group Focus	Literacy/ Reading Instruction	24/ 36%
20 – Use of Basal	Resources	22/33%
9 – Two Hours of Reading	Literacy/Reading Instruction	21/ 32%

N = 66 Kindergarten Classrooms

classrooms where there was no evidence of these descriptors ranged from 32 to 49 percent.

In the first grade classrooms observed, only three of the checklist items had a rating of "0" (no evidence) in more than 30 per cent of the classrooms. One of the checklist items had a rating of "0" in more than 50 percent of the entire sample of first grade classrooms observed (Table 22).

Table 22

Top Items with a Score of "0" in First Grade Classes

Item	Domain	Frequency/ Percent
6 – Student Work	Environment/ Print	34/ 57%
17 – No Worksheets	Literacy/ Reading Instruction	26/ 43%
16 – Center Work	Literacy/ Reading Instruction	22/ 37%

N = 60 First Grade Classrooms

Checklist items that received the greatest number and percentage of scores with a "0" were from two of the four instrument domains. Like the kindergarten classrooms, the majority of the instrument items were from the literacy/ reading instruction domain. The domains not

represented were resources and planning/ preparation. The percentage of first grade classrooms where there was no evidence of these descriptors ranged from 37 to 57 percent.

In the second grade classrooms observed, six of the checklist items had a rating of "0" (no evidence) in more than 30 per cent of the classrooms. Two of these checklist items had a rating of "0" in more than 50 percent of the second grade classrooms observed (Table 23).

Table 23

Top Items with a Score of "0" in Second Grade Classes

Item	Domain	Frequency/ Percent
22 – Big Books	Resources	40/ 71%
6 – Student Work	Environment/ Print	32/ 57%
15 – Centers	Literacy/ Reading Instruction	25/ 45%
16 – Center Work	Literacy/ Reading Instruction	21/ 38%
10 – Shared Reading	Literacy/ Reading Instruction	20/ 36%
26 – Reading Levels	Planning/ Preparation	18/ 32%

N = 56 Second Grade Classrooms

Checklist items that received the greatest number and percentage of scores with a "0" were from all four of the instrument domains. Three of the six items were from the literacy/ reading instruction domain. The percentage of classrooms where there was no evidence of these descriptors ranged from 32 to 71 percent.

Chapter Summary

This study was guided by two research questions. Research question one, can the implementation of a school or division reading program be assessed in a valid, reliable, and efficient way, was addressed through the development and use of a primary reading classroom

observation checklist. In this chapter, the data gathered from the classroom observations that were conducted using the observation instrument were reported by instrument rating for each of the checklist items as well as by school type (Title One or non-Title One) and by grade level designation (kindergarten, first, or second grade classroom). In an effort to address research question two, what accounts for the greatest variance in the implementation of a district's primary literacy program, the instrument items that had the highest percentages of "clearly evident" and "no evidence" ratings were reported both by school type as well as by grade level designation.

There are several instrument items that received the greatest number and percentage scores of "2" (clearly evident) across school type and grade level designation (Table 24) in more than 70% of the classrooms observed.

Table 24

Summary of Top Scoring "Clearly Evident" Instrument Items

Top Items All Classrooms	Top Items Title One	Top Items Non-Title One	Top Items K	Top Items 1st	Top Items Second
8	23	8	22	25	23
25	8	25	1	23	18
23	25	9	8	9	20
1	1	19	19	1	25
19	19	1	25	20	9
9					

Instrument items 8 (Breakthrough to Literacy program is in use) and 19 (Breakthrough to Literacy books) appear as a high scoring instrument items in all possible categories (note: since this program is found only in the kindergarten classrooms of the division used as the sample site, these instrument items were not rated on the checklist in the first and second grade classrooms). Checklist item 25 (all teachers assigned to the classroom work directly with students during all components of reading instruction) also appears in all six of the possible categories. Other checklist items that appear in multiple categories include item one (well-defined areas for whole group instruction and small group work), which had a greater than 70% frequency in all categories but second grade classrooms as well as checklist items 2 (all teachers have lesson plans) which appear in all but the non-Title One and kindergarten categories.

Similarly, the instrument items that received the greatest number and percentage scores of "0" (no evidence) across school type and grade level designation in more than 30% of the classrooms observed (Table 25) have some commonalities.

Table 25

Summary of Top Scoring "No Evidence" Instrument Items

Top Items All Classrooms	Top Items Title One	Top Items Non-Title One	Top Items K	Top Items 1st	Top Items 2nd
6	6	6	6	6	22
15	15	26	26	17	6
16	22	16	15	16	15
26		15	16		16
17		17	11		10
22			14		26

20
9

Checklist item six (students' original, current work supporting reading/writing is posted) appears in all six categories as having "no evidence". Checklist items 15 (students are working in literacy centers during teacher-directed small group instruction) and 16 (literacy center work is open-ended) appear in five of the six "no evidence" categories. Item 15 is not in the first grade category, and item 16 does not appear in the Title One category. Checklist item 26 (grouping for guided reading is based on reading level) appears in all categories but non-Title One and second grade. Checklist items 17 (little or no use of worksheets) and 22 (use of big books) appear in all categories but Title One (item 17) and non-Title and first grade (item 22). Title One and first grade classrooms had the fewest number of checklist items that received scores of "0" in more than 30 percent of the classrooms observed.

The number of checklist items that were clearly evident in at least 70 percent of the sampled classrooms across grade levels and school types was almost consistently five (20 percent). However, the number of checklist items that were not evident in at least 30 percent of the classrooms ranged from three in the Title One and first grade classroom samples to eight in the kindergarten classrooms.

The purpose of this chapter was to report the results of the development and testing of a formative reading program implementation assessment model. Each of the checklist items represents a specific independent variable that was reported by percent in an effort to identify the degree of implementation of the reading program components. A discussion of these results along with conclusions and recommendations for future study are included in Chapter V.

CHAPTER V

SUMMARY OF FINDINGS

The purpose of this study was to develop and test a classroom observation inventory as a formative reading program implementation tool. This chapter provides a summary and a discussion of the results of the study. It also presents recommendations for practice and for further research.

Review

Recent research in effective reading instruction (e.g. Mitchell & Wile, 2002; Taylor, Pearson, Peterson, & Rodriguez, 2003) as well as federal legislation (No Child Left Behind Act of 2001) have had a significant impact on early literacy instruction in schools across the nation. Previous studies also have identified specific reading program attributes and their relationship to increased student achievement by isolating variables of effective reading programs in schools and classrooms (e.g. Allington & Johnson, 2000; Pressley, Rankin, & Yokoi, 1996). The checklist developed for this study used major components of a research-based school division elementary reading plan as a means of answering research question one (How can the implementation of a school or district reading program be assessed in a valid, reliable, and efficient way?). The checklist items were drawn directly from the plan itself as well as from training sessions the division held on the implementation of the plan. The checklist was divided into four domains and after testing for content and face validity, the final version used in this study had 26 items.

For this study, sixteen observers selected from the school division central office staff were trained in scoring the classroom observation instrument using a structured training agenda. After a training session, schools that were selected using random stratified sampling were

alphabetically assigned to pairs of observers. The eight pairs of observers were given a two-week window to conduct their observations, ultimately in 182 primary classrooms in fifteen elementary schools. For each of the 182 classrooms that were observed using the checklist, the grade level of the classroom and the type of school (Title I or non-Title I) were coded. A rating of either "2" (clearly evident), "1" (somewhat evident), or "0" (not evident) was then assigned to each of the 26 items on the instrument.

The frequency and percent of the sample of observed classrooms that received each of the ratings were then calculated and organized by the four instrument domains according to school type and grade level configuration. In an effort to answer research question two (What accounts for the greatest variance in the implementation of a division's primary literacy program?), the instrument items that received the greatest number and percentage scores of "2" (clearly evident) and "0" (no evidence) were also reported for the entire sample as well by grade level and by school type.

Findings

This study was centered around the development, testing and use of a classroom observation instrument that would quantitatively determine the degree to which a primary literacy program was being implemented in a school division. A second purpose of this study was to determine if there was any variance in the way the program was implemented across a particular school division.

Finding One

The primary literacy program was being fully or partially implemented in at least 70 percent of the classrooms observed as part of this study. For example, 133 classrooms (or 73%) clearly had at least two hours of daily reading instruction. In 144 (79%) of the classrooms that

were observed, all of the adults in the classroom were working directly with students. Lesson plans for reading instruction were clearly available in 142 (78%) of the classrooms. The division's adopted basal reading resource was clearly or somewhat evident in 143 (78%) of the classrooms, and the leveled texts were evident in 141 (or 73%) of the classrooms. In the kindergarten classrooms that were observed as part of this study, 61 classes (or 94%) had at least some evidence of using the division adopted Breakthrough to Literacy print resources, as well as using the Breakthrough to Literacy computer program.

Finding Two

The data showed that the indicator related to the posting of student work (instrument item six) was not being implemented. In the entire sample, the instrument item related to this literacy program indicator was clearly evident in only 17 per cent of the kindergarten through second grade classrooms observed. More specifically, the posting of student work was present in 13 percent of the Title One classrooms. In the non-Title One classrooms, student work was somewhat more visible, being clearly evident in 21 percent of these classrooms. Student work was more visible in first grade classrooms than in the kindergarten or second grades; however, the differences between the three grade levels are small. Student work was clearly evident in 15 percent of the kindergarten classrooms, 20 percent of the first grade classrooms, and in 16 percent of the second grade classes.

Finding Three

In addition to student work, the data also showed a low level of implementation on the literacy program components pertaining to the use of literacy centers in classrooms. For this study, program elements that were evident in at least 70 percent of the classrooms was considered acceptable. Instrument items 15, 16, and 17 focused on the presence of literacy

centers in the classrooms as well as the type of work students were doing while at the centers. Across all grade levels and school types, the level of clearly evident implementation ranged from 30 to 43 percent on these three items. While 43 percent of the classrooms had literacy centers that were clearly evident, only 37 percent of the classrooms had students doing authentic work in the centers that was not dependent on worksheets, and only 30 percent of the work that students were doing was open-ended in nature.

Finding Four

The data showed that "environment/ print" and "planning and preparation" had the highest levels of implementation across grade levels and school types. There were 31 literacy program indicators that were top scoring "clearly evident" (Table 24). Of these top scoring "clearly evident" indicators, 19 (or 66 percent) were from "environment/ print" and "planning/ preparation". Nine of the19 literacy program elements were from "environment/ print" and 10 of the items were from "planning/ preparation". These results indicate that the teachers that were observed as part of this study had developed literacy lesson plans and had created classroom environments that were conducive to effective literacy instruction.

Finding Five

The data showed that "literacy/ reading instruction" had the lowest level of implementation. There were 30 instrument items that were top scoring "no evidence" instrument items (Table 25). Of these top scoring "no evidence" items, 17 (or 56 percent) that had "no evidence" across school types and grade levels came from "literacy/ reading instruction". Several previous studies (Barber, 2002; Curry, 1999; Foertsch, 2000; Guthrie, 2000; Mitchell, 2002; Pressley, 1996; Pressley, 2001; Pressley, 2003; & Taylor, 1996) identified as an essential element of effective primary literacy instruction – daily direct instruction, or the time spent in

whole and small groups developing reading skills and strategies. This crucial instructional element was either missing or only somewhat evident in 44 percent of the classrooms that were observed as part of this study.

Finding Six

After disaggregating the observation data by grade level (kindergarten, first and second grade), the data showed that each domain of an effective reading program was implemented to a different degree. For example, when just looking at grade level data (and omitting school type), "resources" and "planning/ preparation" had the highest levels of implementation. In addition, each of the three specific grades levels had five literacy program elements that were clearly evident in more than 70 percent of the classrooms observed. In the first grade classrooms, these program elements were from "environment/ print", "literacy/ reading instruction", "resources", and "planning/ preparation". In the second grade classrooms, the items were from "literacy/ reading instruction", "resources", and "planning/ preparation". In the kindergarten classrooms, the highest scoring items were from "environment/ print", "resources", and "planning/ preparation". The domain not represented as high scoring in the kindergarten classrooms was "literacy/ reading instruction".

Finding Seven

The first grade classrooms had a higher overall level of implementation than their kindergarten and second grade counterparts. This is evidenced in this study through the three program elements in the first grade classrooms that received high (greater than 30 percent) "no evidence" scores. This contrasts with the eight literacy program elements in the kindergarten classrooms that fell into this category, and the six literacy program elements in second grade classrooms that received high "no evidence" scores. Besides the high numbers of "no evidence"

program elements in the kindergarten and second grade classrooms, five (or 62%) of the literacy program characteristics with low levels of implementation in the kindergarten classrooms were from “literacy/ reading instruction”, and three of the literacy program elements with low levels of implementation in second grade classrooms were also from “literacy/ reading instruction”.

Finding Eight

Title One classrooms had higher percentages of implementation than their non-Title One counterparts. The data showed that the number of instrument items that received a “clearly evident” score in more than 70 percent of the Title One and non-Title One classrooms was exactly the same (five items). However, on 20 of the 26 items (or 77 percent), the clearly evident scores were higher in Title One classrooms than their non-Title One counterparts. In addition, 17 of the 26 mean instrument items were higher in Title One schools. For example, 57 Title One classrooms (62%) were using leveled books to deliver reading instruction while only 43 non-Title One classrooms (48%) were doing the same. The mean raw score for this item in Title One schools was 1.46 compared with 1.09 in non-Title One classrooms. In these same non-Title One classrooms, 40 (or 44% with a mean raw score of 1.27) had a word wall posted while 53 Title One classrooms (58% with a mean raw score of 1.45) had a word wall. Similarly, in the Title One classrooms 49 (53% with a mean raw score of 1.30) of the teachers had developed lesson plans based on the instructional reading levels of their students, while in the non-Title One classrooms 28 teachers (31% or a mean raw score of .83) had used reading levels in their lesson plans.

Finding Nine

The data also showed that the implementation of a division’s reading program can be assessed in a valid and reliable way. The instrument that was used to determine the level of

implementation of the primary reading program had a content validity of greater than 80 percent (Appendix F). In addition, the instrument was piloted to improve the content and face validity of it. Inter-rater reliability was used to ensure consistency of scoring with the instrument used in this study to obtain agreement between the pairs of observers on the ratings for each of the instrument items that described components of the primary literacy program. Observation pairs observed the first classroom together, and they did not proceed on their own until they had at least 85 percent agreement (Appendix H). Prior to the actual observations, the 16 classroom observers attended a two-hour training session which focused on obtaining a high degree of inter-rater reliability by using a well-defined set of guidelines for scoring the instrument that measured the degree of implementation of the reading program. The observations occurred during a short window of time to reduce some threats to internal validity that could occur primary through instrument use history and in an effort to ensure that the environment in which the study happened in all schools that were selected was the same. In addition, the short window of time reduced the threat of maturation in that once administration of the instrument began, it potentially was no longer measuring a neutral environment.

Conclusions

The findings from the use of this observation instrument show that first grade classrooms and the primary classrooms in Title One schools have the strongest degree of implementation of the components of a research-based primary literacy program in this particular school division. In contrast, the findings also indicate that non-Title One schools as well as the kindergarten and second grade classrooms had a greater degree of no evidence of implementing the plan's components as identified through the "no evidence" scores on the classroom observation instrument used in this study. Of the four sections of the classroom observation instrument,

"environment/ print" and "planning/ preparation" had the highest level of implementation across school type and grade level while "direct literacy/ reading instruction" had the lowest level of implementation across school type and grade levels. This particular finding contradicts previous research which found that the literacy program elements found in "direct literacy/ reading instruction" are the most crucial, yet in the classrooms observed in this study they were the ones that were not being done.

Recommendations for Practice

The purpose of assessing the impact of any reading program is to determine the effectiveness of the program, use the results to improve instruction, and ultimately increase student achievement. This study was designed to use a formative or process program implementation model to answer two basic questions:

How can the implementation of a school or division reading program be assessed in a valid, reliable, and efficient way?

What accounts for the variance in the implementation of the primary reading program across school types and grade levels?

In this study's sample, finding one indicated that the majority of the classrooms that were observed were either fully or partially implementing at least 70 percent of the literacy program components. As a result of this finding, it is recommended that this particular school division develop a short and long-range plan for monitoring the implementation of its literacy program. The school and central office administrators in this division must ask the question, "Is a 70 percent implementation rate good enough?" In this particular sample of half of a school division's primary classrooms, the number of classrooms not implementing the division's reading plan components would be 108 classrooms. If the division's average primary class size is

15 students, then ultimately 1,620 students could potentially be impacted by the lack of evidence of particular instrument items.

Findings two and three from this study showed that particular checklist items had large percentages of non-implementation. In particular, this included the checklist items related to the posting of current student work in the classroom and the use of literacy centers and the type of work students were engaged in while at the centers. The lack of student work in teachers' classrooms is a communication, expectation and monitoring issue on the part of a building administrator. However, the lack of literacy centers and the quality of the activities that were observed in the centers is not only a communication, expectation, and monitoring issue, but also a staff development issue. For one, basic sessions on the nature and purpose of using literacy centers must be conducted. In addition, follow-up sessions need to be conducted on developing appropriate activities for students to use while they are working at literacy centers. And like the previous instrument item on student work, this also becomes a communication, expectation, and monitoring issue for the building administrator. As a result of these two findings, it is recommended that building administrators engage primary teachers in professional development activities surrounding the reading program elements related to these specific instrument items.

Findings four, five and six from this study all found a high level of implementation of the "environment/ print" and "planning/ preparation" literacy program elements and a low level of implementation of "literacy/ reading instruction" across grade levels and school types. The findings from this study also showed a disconnection between the abundance of previous research that found this specific area of a literacy program as crucial and the implementation of it in this particular school division. As a result of these three findings, it is recommended that building administrators in this division first focus their professional development efforts on the

implementation of the instrument item components that are found under this particular literacy program area. Specifically, it is recommended that the professional development focus on the teaching of shared and guided reading and word study and the implementation of literacy centers in each classroom. The last portion of this recommendation would encompass the recommendation related to findings two and three as well. The building level reading specialist would also play a critical role in the demonstration, modeling, and coaching of these particular reading program components. Previous researchers (i.e. Fisher, 1999; Mehall, 2001; Musenthal, 2003) have also noted the value of having building-level literacy leaders in roles that promote, support, and champion classroom reading instruction. As a result of this research and the data from this study, it is also recommended that divisions and schools develop and implement a plan for actively engaging their building level experts (the reading specialist) in coaching, modeling, and supporting teachers in implementing specific reading plan components related to these findings and this recommendation for practice.

Findings seven and eight from this study indicated that the Title One and first grade classrooms had a higher level of implementation of the reading plan's components than their non-Title One, kindergarten and second grade counterparts. The implications of these results for this particular school division are obvious. As a result of these particular findings, it is recommended that principals and appropriate division-level personnel develop a short and long-range plan for the improvement of specific literacy instruction components in all primary classrooms, and that the division more closely monitor the implementation of its reading program. It is also recommended that, in an effort to avoid similar voids, school administrators and related central office staff from other school divisions modify the checklist used in this study as a means of determining whether particular elements of their reading programs are being

implemented, or to determine whether particular schools or grade levels are implementing the program components more than others and whether this improved level of implementation results in higher reading scores. Furthermore, based on these particular findings, it is also recommended that a short-term and long-range professional development plan be developed to ensure that the teachers in the division have the capacity to implement the reading plan's various components. Initial training is needed for veteran staff, and it also needs to be built into the teacher induction program as veteran teachers leave and new staff is hired. In addition to this short- term training, a long-range menu of professional development offerings needs to be developed tailored to the grade levels or schools where the literacy program indicators had high percentages of "no evidence".

Finding nine showed that the implementation of a division's primary literacy program can be assessed in a valid, reliable, and efficient way. Ultimately, a specific instructional program is implemented in a school or school division to produce a desired result: increased student achievement. When a research-based program is clearly implemented across grade levels and schools, that result will be realized. As a result of this finding, it is recommended that building or division administrators implement a formative study using the instrument developed for this study that would determine the program's level of implementation. It is also recommended that once the areas of non-implementation are identified, subsequent actions would be taken by both building and division staff to see that the program is fully implemented rather than at the 70 percent level that was identified as acceptable in this study.

Recommendations for Further Study

There are several recommendations for further study that could occur as an outgrowth of this study:

1. Conduct a study that extends the classroom observation instrument to include grades three through five. Such a comprehensive literacy study of kindergarten through fifth grade would provide a comprehensive picture of a reading program in an entire elementary building or elementary portion of a school division.
2. Extend the classroom observation instrument to include specific components of the instrument items in an attempt to assess the quality of the implementation of those items.
3. Investigate the reason why more Title I schools tended to implement more reading program elements than non-Title One schools through the development of a structured interview protocol used with both teachers, reading specialists and principals.
4. Investigate why first grade classrooms tended to implement more reading program elements than kindergarten and second grade classrooms through the development of a structured interview protocol used with both teachers, reading specialists and principals.
5. Identify schools whose students consistently produce high reading test scores at the end of third grade. Use a version of the quantitative checklist from this study to determine if there are particular independent variables or domains in the kindergarten, first and second grade classrooms of these specific schools that are stronger than others.
6. Conduct a longitudinal study using this instrument in conjunction with a series of related professional development activities to determine the relationship between the professional development and the subsequent implementation of the components of effective reading instruction.
7. Investigate the relationships between the results of a similar observation checklist study and subsequent student achievement in reading. Determine if there is a relationship

between “clearly evident” classrooms and the subsequent student success. Determine if “clearly evident” classrooms out perform “somewhat evident” classrooms.

8. Examine why there was a higher level of implementation of the division’s reading program in first grade, and a subsequent break down in implementation in kindergarten and second grade.
9. Investigate the relationship between level of primary reading program implementation in accredited and non-accredited schools.
10. Determine if there is a relationship between school level administrative support and the level of implementation of the primary reading program.
11. Extend the amount of classroom observation time and investigate the quality of the implementation of various primary reading program components noted on the checklist used for this study.

Limitations

Just as there are a variety of models for determining the degree of program implementation as well as a program’s effectiveness, there are a variety of problems associated with conducting a study of this nature and more specifically, conducting a study that includes collecting data about teaching. Often teachers adjust their behaviors when being observed, particularly if they know why they are being observed and thus the results of the study are altered. In other words, were the behaviors that were observed typical and representative, or were they non-representative.

The findings from this study are also limited to the school division where it was conducted and to the research-based reading plan that was used to develop the classroom observation instrument that was used to conduct it. A similar instrument that was used, even in a

similar district, could produce very different results. Also, the instrument was designed to determine whether particular components of a research-based primary reading program were being implemented in the classrooms of a school division. The instrument was not designed to assess the quality of that implementation.

While this study's population was representative of the school district as a whole (geographically and economically), the study did not identify specific schools by name or through coding of the observation sheets. As a result, this study does not provide comparative data beyond that of school type and grade level. For example, this study does not provide a means of comparing high achieving schools to low achieving ones. Although the population of schools chosen for this study was geographically representative of the district as a whole, there was no way of determining where the reading program was better implemented in one particular area of the district over another.

In addition, 16 observers conducted this study during a two-week window of time. As a result, the study is limited in that this was a "snapshot" of primary reading instruction in a particular school division and was not a series of observations conducted over an entire school year. And, although this study incorporated a training session for the raters as well as a test of inter-rater reliability, the skill level and consistency of using multiple observers to conduct this study is also a limitation. However, the quality and quantity of the data this study provided as well and the multitude of subsequent actions that could come as a result of it more than compensate for the limitations of this or a similar study.

Research Reflections

Conducting this study has given this researcher a great deal of understanding of the formal quantitative research process. Although the idea of gathering data from almost 200

classrooms and tabulating the subsequent results seemed very daunting, the actually management of the data was actually exciting. I was apprehensive about involving sixteen other people in the study, especially since the ultimate success of the study was dependent on them conducting the observations according to the guidelines that were established in the training session, using the instrument to gather the data, and conduct the observations within a two-week window of time. Overall, the entire process went off without any major stumbling blocks.

Nationally as well as in the Commonwealth of Virginia, the focus on school division accountability as defined by student achievement in reading continues to escalate. And although discussion continues about specific assessment benchmarks for subgroups of students, one of the chief components of the No Child Left Behind Act, all students reading on grade level by 2014, remains. This high degree of accountability escalates the need for school administrators at all levels to either begin or refocus their efforts on monitoring their literacy programs. The results of this study, as well as those that could be replicated from it will assist administrators in determining the degree to which their literacy programs are operationalized and thus increase student achievement.

References

Allington, R., & Johnston, P. H. (2000). *What do we know about effective fourth-grade teachers and their classrooms?* Retrieved February 8, 2004, from The National Center on English Learning & Achievement Web site: http://cela.albany.edu

Barber, M. S. (2002). The influence of the Alabama Reading Initiative on the culture of three inner-city elementary schools in the state of Alabama. *Dissertation Abstracts International, 64*(01A), 95.

Bonifay, S. (1993). An evaluation and critique of a middle school reading program: A case study. *Dissertations Abstract International, 54*(12), 4312.

Brown, R., Pressley, M., Van Meter, P., & Schuder, T. (1996). A quasi-experimental validation of transactional strategies instruction with low-achieving second-grade readers. *Journal of Educational Psychology, 88*(1), 18-37.

Critical issue: Monitoring the school literacy program. Retrieved February 13, 2004 from the North Central Regional Education Laboratory Web site: http://www.ncrel.org

Curry, J. & Ayskowski, G., (1999). Summer opportunity to accelerate reading: Evaluation of a balanced reading program for low achieving students. *ERS Spectrum, 17*(3), 3-10.

Dressel, P. L. (1976). *Handbook of academic evaluation.* San Francisco: Josey Bass.

Dunkin, M. J. & Biddle, B. J. (1974). *The Study of Teaching.* New York: Holt.

Farr, R., & Pritchard, R. (1989). Evaluation of district reading programs. In S. Wepner, J. Feeley, & D. Strickland (Eds.), *The administration and supervision of reading programs* (pp. 162-176). New York: Teachers College Press.

Farr, R., & Greene, B. (1999). *A guide for evaluating a reading or language arts program.* Bloomington: Indiana University, Center for Innovation in Assessment.

Fink, A. (1995). *Evaluation for education and psychology*. London: Sage.

Fisher, C., & Adler, M. A. (1999). *Early reading programs in high-poverty schools: Emerald Elementary bests the odds.* Ann Arbor: Center for the Improvement of Early Reading Achievement.

Fisher, R., Lewis, M., & Davis, B. (2000). Progress and performance in National Literacy Strategy classrooms. *Journal of Research in Reading, 23*(3), 256-266.

Fitzpatrick, J. L. (2004). *Program evaluation: Alternative approaches and practical guidelines* (3rd ed.). Boston: Pearson.

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

Foertsch, M. (2000). *A study of reading practices, instruction, and achievement in District 31 Schools.* Oak Brook, IL: North Central Regional Educational Laboratory.

Gibton, D. (2002). Evaluating serendipity. *Studies in educational evaluation,* 28, p. 189-198.

Guba, E. G. & Lincoln, Y. S. (1981). *Effective evaluation.* San Francisco: Jossey-Bass.

Guthrie, J. T., Schafer, W. D., Von Secker, C., & Alban, T. (2000). Contributions of instructional practices to reading achievement in a statewide improvement program. *Journal of Educational Research, 93*(4), 211-225.

Identifying and implementing educational practices supported by rigorous evidence: A user friendly guide (2004). Washington, DC: Coalition for Evidence-Based Policy.

Inman, D., Marlow, L., & Barron, B. (2004). Evaluation of a standards-based supplemental program in reading. *Reading Improvement, 41*(3), 179-186.

Learning First Alliance (1998). Every child reading: An action plan of the Learning First Alliance. *American Education*, *22*(1), p. 52 – 63.

Levine, T. (2002). Stability and change in curriculum evaluation. *Studies in educational evaluation, 28,* p. 1-33.

Linn, R. L. & Gronlund, N. E. (1998). *Measurement and assessment in teaching*. Columbus, OH, Merrill.

Lomax, R. G. (2004). Whither the future of quantitative literacy research? *Reading Research Quarterly, 39*(1), p. 107-112.

McNamara, C. (2003). Basic guide to program evaluation. *Retrieved February 18, 2004 from the Web site: http://www.mapnp.org/library/evaluatn/fnl_eval.htm*

Mehall, L. (2001). Examining the impact of the Illinois quality assurance review process: Changes in curriculum, instruction, and assessment in reading in two middle schools. *Dissertations Abstract International, 63*(09), 3074.

Mitchell, S. & Wile, N. (2002). The *2001 literacy program evaluation: A report of the evaluation of programs in elementary and middle schools.* Portland, OR: Portland Public Schools.

Mosenthal, J., Lipson, M., Mekkelsen, J., Russ, B., & Sortino, S. (2003). *Elementary schools* w*here students succeed in reading*. Providence: The Education Alliance.

National Center for Educational Statistics. (1996). NAEP 1994 reading report card for the nation and the states. (NCES Report No. 96045). Washington, DC: Office of Educational Research and Improvement, U.S. Department of Education.

National Institute of Child Health and Human Development. (2000). *Report of the National Reading Panel. Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction* (NIH Publication No. 00-4769). Washington, DC: U.S. Government Printing Office.

Newmann, F. M., Smith, B., Allensworth, E., & Bryk, A. S. (2001). Instructional program coherence: What it is and why it should guide school improvement policy. *Educational Evaluation and Policy Analysis, 23*(4), 297 – 321.

No Child Left Behind Act of 2001, 20 U.S.C. § 6801 (2002).

O'Neal, M. R., Snyder, S. W., & Spor, M. W. (2001). *Evaluation of year 2 of the Alabama Reading Initiative.* Paper presented at the meeting of the Mid-South Educational Research Association, Little Rock, AR.

Pressley, M., Rankin, J., & Yokoi, L. (1996). A survey of instructional practices of primary teachers nominated as effective in promoting literacy. *The Elementary School Journal, 96*(4), 363-384.

Pressley, M., Wharton-McDonald, R., Allington, R., Block, C., Morrow, L., Tracey, D., et al. (2001). A study of effective first-grade literacy instruction. *Scientific Studies of Reading, 5*(1), 35-38.

Program evaluation standards: How to assess evaluations of educational programs (2nd ed.). (1994). London: Sage.

Reynolds, D. (1998). Schooling for literacy: A review of research on teacher effectiveness and school effectiveness and its implications for contemporary educational policies. *Educational Review, 50*(2), 147-1162.

Smith, N. L. (1994). Evaluation models and approaches. In *The International Encyclopedia of Education* (pp. 2101-2109).

Snow, C. E., Burns, C., & Griffin, P. (Eds.) (1998). *Preventing reading difficulties in young children: Report of the Committee on the Prevention of Reading Difficulties in Young Children.* Washington, DC: National Academy Press.

Stufflebeam, D. L. (1983). The CIPP model for program evaluation. In G. F. Madus, M. Scriven, & D. L. Stufflebeam (Eds.), *Evaluation models: Viewpoints on educational and human services evaluation.* Boston: Kluwer-Nijhoff.

Taylor, B.M. & Pearson, P. D. (2001). The CIERA School Change Project: Translating research on effective reading instruction and school reform into practice in high-poverty elementary schools. In C. M. Roller (Ed.), *Learning to teach reading: Setting the research agenda* (pp. 180 – 189). Newark, DE: International Reading Association.

Taylor, B. M. & Pearson, P. D. (2000). *The CIERA school change classroom observation scheme*. Minneapolis, MN: University of Minnesota.

Taylor, B. M., Pearson, P. D., Clark, K. F., & Walpole, S. (1999). Effective schools/ accomplished teachers. *The Reading Teacher, 53*(10), 23-30.

Taylor, B. M., Pearson, P. D., Peterson, D. S., & Rodriguez, M. C. (2003). Reading growth in high poverty classrooms: The influence of teacher practices that encourage cognitive engagement in literacy learning. *Elementary School Journal 104*(1), 13-27.

Taylor, B. M., Pearson, P. D., Peterson, D. S., & Rogrigues, M. C. (2005). The CIERA school change framework: An evidence-based approach to professional development and school reading improvement. *Reading Research Quarterly, 40* (1), 40 – 68.

Taylor, B. M., Peterson, D., Rodriguez, M. C., & Pearson, P. D. (2002). *The CIERA school change project: Supporting schools as they implement home-grown reading reform.* Ann Arbor: Center for the Improvement of Early Reading Achievement.

Taylor, B. M., Pressley, M., & Pearson, P. D. (2002). Research-supported characteristics of teachers and schools that promote reading achievement. In B. Taylor & P. D. Pearson (Eds.) *Teaching reading: Effective schools, accomplished teachers* (pp. 362-373). New York: Lawrence Erlbaum.

Tuckman, B. W. (1985). *Evaluating instructional programs*. Newton, MA: Allyn and Bacon.

Walkup, P. (1997). Mandated school reform in Illinois: The case of elementary reading. *Dissertations Abstract International, 58*(12), 4602.

Worthen, B. R. & Sanders, J. R. (1987). *Educational evaluation: Alternative approaches and practical guidelines.* New York: Longman.

Appendix A

KINDERGARTEN REQUIRED DAILY SEGMENTS

WHOLE GROUP INSTRUCTION: Shared Reading/Oral Language Development (30-60 minutes)

Direct instruction using big books in:

- Building background knowledge
- Phonological awareness
- Concepts of print
- Alphabet recognition/principles
- Letter-sound knowledge
- Listening comprehension strategies and skills
- Developing sight vocabulary (word recognition skills)
- Story elements

SMALL GROUP INSTRUCTION/CENTERS: (50-60 minutes)
(ongoing assessment guides group formation)

- Three teacher-directed flexible guided reading groups using big books and adopted text
- Direct SOL-based instruction in:
  - Phonemic awareness
  - Letter recognition
  - Letter/sound recognition
  - Concepts of print
  - Listening comprehension skills
  - Word recognition
- Centers
  Independent reading using classroom libraries
  Skill and strategy practice with authentic activities
  Use of Word Wall for word study and center work
  Use of Breakthrough to Literacy for phonemic awareness practice

Appendix B

FIRST GRADE REQUIRED DAILY SEGMENTS

WHOLE GROUP INSTRUCTION: Shared Reading/Oral Language Development/Mini - Lesson (up to 45 minutes)

Direct instruction using big books and adopted basal text in:

- Building background knowledge
- Phonological awareness and letter/sound knowledge
- Concepts of print and text structure
- Comprehension strategies and skills
- Story elements
- Word study using word wall
- Developing sight vocabulary (word recognition skills)

SMALL GROUP INSTRUCTION (60-90 minutes)
(ongoing assessment guides group formation)

- Three teacher-directed flexible guided reading groups
- Direct SOL-based instruction using leveled text in:
  - Phonemic awareness
  - Comprehension skills and strategies
  - Vocabulary development
  - Word study (word recognition skills) using word wall and other materials
- Literacy/Extension centers
- Reinforcement learning activities
  - Independent reading using classroom libraries
  - Skill and strategy practice with authentic activities

Appendix C

SECOND GRADE REQUIRED DAILY SEGMENTS

WHOLE GROUP INSTRUCTION: Shared Reading (20-30 minutes)

Direct instruction using big books and adopted basal text in:

- Oral language development
- Phonological awareness and letter/sound knowledge
- Building background knowledge
- Text features and genre
- Comprehension strategies
- Word Study
- Story Elements
- Sight vocabulary development (word recognition skills)

THREE TEACHER-DIRECTED FLEXIBLE GUIDED READING GROUPS (60 - 90 minutes) (ongoing assessment guides group formation)

Direct instruction/modeling using leveled text:

- Phonological awareness and spelling patterns
- Comprehension strategies and skills
- Vocabulary development (sight and meaning)
- Fluency building
- Word Study
- Literacy/Extension centers
- Reinforcement learning activities
    - Independent reading using classroom libraries
    - Skill and strategy practice with authentic activities
    - Use of Word Wall for word study and center work
- Varied expository text structures (social studies/science)/genres

APPENDIX D
Initial K-2 Literacy Checklist

Grade: K 1 2 **Title I School: Yes No**

Key: 2 = Clearly Evident 1 = Some Evidence 0 = No Evidence

	2	1	0
Environment/Print			
1. Well-defined areas for whole group instruction and small group work.			
2. Routines for literacy instruction are established and followed.			
3. Literacy environment is clean and organized.			
4. Classroom is print rich.			
5. A word wall, organized alphabetically, of high frequency words and words to represent spelling patterns is present.			
6. Students' original, current work supporting reading/writing is posted.			
7. Classroom library exhibits a variety of reading materials/levels.			
8. Breakthrough to Literacy program is in use (K only)			
Literacy/Reading Instruction **Daily components of reading plan are evident and include:**			
9. 2 to 2 ½ hours of daily communication skills instruction.			
10. Shared Reading.			
11. Guided Reading.			
12. Word Study.			
13. Focus of whole group instruction is the teaching of reading and writing skills/strategies.			
14. Focus of small group instruction is the prompting and modeling of reading and writing skills/strategies.			
15. Students are working in literacy (reading, writing, listening) centers during teacher-directed small group instruction.			
16. Literacy center work is open-ended.			
17. Little or no use of worksheets.			
18. Effective monitoring of student learning is evident.			
19. Student progress in reading is regularly monitored as evidenced by complete K-2 literacy assessment records.			
Resources (Appropriate texts are used)			
20. Breakthrough to Literacy books (K Only)			
21. Current basal (Harcourt *Signatures*)			
22. Sets of leveled books			
23. Big books			
Planning/Preparation			
24. All teachers assigned to the classroom have individual lesson plans written for the day.			
25. Lesson plans are written for all components of the communication skills block.			
26. All teachers assigned to the classroom work directly with students during all components of reading instruction.			
27. Grouping for guided reading is based on reading level.			
Administrative Support for Literacy			
28. Principal is knowledgeable about the location and role of literacy resource teachers			
29. All literacy support teachers are in classrooms working directly with students and teachers:			
30. Communication Skills Specialist			
31. Literacy Support Teacher			
32. Literacy Support Teacher			

APPENDIX E

CONTENT VALIDATION INSTRUMENT FOR OBSERVATION CHECKLIST

Directions: Circle the number of the appropriate response.

Domains:

(1) Environment/Print – a print rich classroom that includes a classroom library, word wall, and well defined areas for whole and small group instruction.
(2) Daily Direct Instruction in Reading – time spend in whole and small groups developing reading skills and strategies.
(3) Resources – appropriate and a variety of resource materials to deliver reading instruction.
(4) Planning/Preparation – plans are developed and used that meet the literacy needs of all students.
(5) Administrative Support – principal and literacy leaders promote, support, and champion reading instruction.

Association Ratings: 1 = very weak, 2 = weak, 3 = strong, 4 = very strong

(For any items you rate as 1 or 2 for association, please write your suggestions for improvement directly on this page.)

Observation Instrument Statement	Domain					Association			
1. Shared Reading	1	2	3	4	5	1	2	3	4
2. Classroom is print rich	1	2	3	4	5	1	2	3	4
3. Routines for literacy instruction are established and followed	1	2	3	4	5	1	2	3	4
4. At least two hours of literacy instruction	1	2	3	4	5	1	2	3	4
5. Literacy work is open-ended	1	2	3	4	5	1	2	3	4
6. Current basal is used	1	2	3	4	5	1	2	3	4
7. All teachers have individual lesson plans written for the day	1	2	3	4	5	1	2	3	4
8. Literacy environment is clean and organized	1	2	3	4	5	1	2	3	4
9. Well defined areas for whole and small group work	1	2	3	4	5	1	2	3	4
10. Grouping for reading is based on reading levels of students	1	2	3	4	5	1	2	3	4
11. Big books	1	2	3	4	5	1	2	3	4
12. Classroom library exhibits a variety of reading materials	1	2	3	4	5	1	2	3	4
13. Focus of small group instruction is the prompting and modeling of reading skills/strategies	1	2	3	4	5	1	2	3	4
14. Principal is knowledgeable about the role of literacy teachers	1	2	3	4	5	1	2	3	4
15. A word wall is present	1	2	3	4	5	1	2	3	4
16. Word study instruction	1	2	3	4	5	1	2	3	4
17. Students original, current work supporting reading/writing is posted	1	2	3	4	5	1	2	3	4
18. Little or no worksheet use	1	2	3	4	5	1	2	3	4
19. Effective monitoring of student learning is evident	1	2	3	4	5	1	2	3	4
20. Guided reading	1	2	3	4	5	1	2	3	4
21. Students are working in literacy centers during small group reading instruction	1	2	3	4	5	1	2	3	4
22. Student progress in reading is regularly monitored	1	2	3	4	5	1	2	3	4
23. Sets of leveled texts	1	2	3	4	5	1	2	3	4
24. Lesson plans are written for all components of the literacy block of time	1	2	3	4	5	1	2	3	4
25. All adults assigned to the classroom work directly with students	1	2	3	4	5	1	2	3	4

APPENDIX F

CONTENT VALIDATION DATA FOR THE K – 2 LITERACY OBSERVATION CHECKLIST

Table 3
Content Validation of Instrument Used for Classroom Literacy Observations: Classification of Items Into Domains by Experts, August 2004 (N = 25)

		Domains									
	Expected Domain	Environment/ Print		Daily Direct Instruction		Use of Resources		Planning and Preparation		Administrative Support	
Item		n	%	n	%	n	%	n	%	n	%
1	Daily Direct Instruction			25	100						
2	Environment/Print	25	100								
3	Environment/Print	24	96					1	4		
4	Daily Direct Instruction			25	100						
5	Daily Direct Instruction			25	100						
6	Use of Resources					25	100				
7	Planning and Preparation			1	4			23	92	1	4
8	Environment/Print	25	100								
9	Environment/Print	24	96	1	4						
10	Planning and Preparation							25	100		
11	Use of Resources					25	100				
12	Environment/Print	25	100								
13	Daily Direct Instruction			22	88			3	12		
14	Administrator Support									25	100
15	Environment/Print	25	100								
16	Daily Direct Instruction			24	96			1	4		
17	Environment/Print	25	100								
18	Daily Direct Instruction			24	96			1	4		
19	Daily Direct Instruction			22	88			2	8	1	4
20	Daily Direct Instruction			24	96			1	4		
21	Daily Direct Instruction			24	96			1	4		
22	Daily Direct Instruction			25	100						
23	Use of Resources					25	100				
24	Planning and Preparation							25	100		
25	Planning and Preparation							25	100		

APPENDIX G

Revised K-2 Literacy Checklist

Grade: K 1 2 **Title I School: Yes No**

Key: 2 = Clearly Evident 1 = Some Evidence 0 = No Evidence

	2	1	0
Environment/Print			
1. Well-defined areas for whole group instruction and small group work.			
2. Routines for literacy instruction are established and followed.			
3. Literacy environment is clean and organized.			
4. Classroom is print rich.			
5. A word wall, organized alphabetically, of high frequency words and words to represent spelling patterns is present.			
6. Students' original, current work supporting reading/writing is posted.			
7. Classroom library exhibits a variety of reading materials/levels.			
8. Breakthrough to Literacy program is in use (K only)			
Literacy/Reading Instruction **Daily components of reading plan are evident and include:**			
9. 2 to 2 ½ hours of daily communication skills instruction.			
10. Shared Reading.			
11. Guided Reading.			
12. Word Study.			
13. Focus of whole group instruction is the teaching of reading and writing skills/strategies.			
14. Focus of small group instruction is the prompting and modeling of reading and writing skills/strategies.			
15. Students are working in literacy (reading, writing, listening) centers during teacher-directed small group instruction.			
16. Literacy center work is open-ended.			
17. Little or no use of worksheets.			
18. Student progress in reading is regularly monitored as evidenced by complete K-2 literacy assessment records.			
Resources (Appropriate texts are used)			
19. Breakthrough to Literacy books (K Only)			
20. Current basal (Harcourt *Signatures*)			
21. Sets of leveled books			
22. Big books			
Planning/Preparation			
23. All teachers assigned to the classroom have individual lesson plans written for the day.			
24. Lesson plans are written for all components of the communication skills block.			
25. All teachers assigned to the classroom work directly with students during all components of reading instruction.			
26. Grouping for guided reading is based on reading level.			

Appendix H

Percentage of Exact Agreement by Observation Pair

Observation Pair	School One	School Two
1	100%	N/A
2	100%	100%
3	96%	96%
4	92%	96%
5	100%	100%
6	100%	100%
7	96%	100%
8	88%	96%

Appendix I

Training Agenda Outline

- Team Member Introductions
- Overview of Purpose of Instrument
- Overview of Development of Instrument
    - Research Base
    - Division Elementary Reading Plan Required Daily Segments
- Overview of Classroom Observation Instrument
- Definition of Instrument Domains
- Definition and Explanation of Individual Instrument Items
- Examples of Artifacts (lesson plan components, portfolio contents, etc.)
- Sample Resource Showcase
- Let's Practice: Video Model
- Collaborative Scoring (in observation pairs)
- Whole Group Scoring and Discussion
- Initial Classroom Observation Guidelines (color-coded sheets)
- Sample Introductory Script/ Personalization of Script
- Questions/ Concerns

Appendix J

Sample Introductory Script

Good morning. Today we are here to observe reading instruction in the kindergarten, first, and second grade classrooms in your building. We will be using a kindergarten through second grade literacy observation checklist developed around our division elementary reading plan as we conduct the observations. We will be visiting about half of the primary classrooms across the city, and the data will be collected and discussed by grade level (kindergarten, first, or second grade) and by school type (Title One and non-Title One). School-specific data will not be reported, nor will it be available at this time. We will need to know the room numbers of all kindergarten, first, and second grade classrooms. Also, we will need to know if there are substitutes in any of these rooms today, as we will not be visiting classrooms where there are substitutes in place of the regular classroom teacher.

Appendix K

Letter of Approval from Local School Division

Norfolk Public Schools

September 10, 2004

Mr. Mark Tavernier
Doctoral candidate
Virginia Tech University

Dear Mr. Tavernier,

On behalf of Norfolk Public Schools (NPS), the Department of Research, Testing and Statistics is pleased to approve your research proposal, *"Formative Elementary Reading Program Assessment: An Interim Tool for Improvement."*

Approval is granted to conduct dissertation research as proposed by Mr. Mark Tavernier, doctoral candidate at Virginia Tech (University), College of Human Resources and Education, Department of Educational Leadership and Policy Studies, in fulfillment of doctoral degree program. Mr. Tavernier is currently employed by Norfolk Public Schools as Senior Coordinator, Communication Skills, Department of Leadership, Capacity and Development.

Permission is granted to use any components from the Norfolk Public Schools Elementary Reading Plan document in the development of the checklist proposed in your study.

Participation in the research is voluntary at all NPS faculty and staff levels. Any participant may withdraw at any time without question or consequence. Any issues or problems related to conducting this research are to be reported immediately to the Dept of Research, Testing & Statistics (RTS).

Approval does not constitute commitment of resources or the endorsement of the study or its findings by the school district or the School Board. All participants are to remain anonymous in data collection and reporting results. Identifiable characteristics or linkage to the identity of any participant is prohibited.

A copy of the results of the study, and any changes to the research as proposed and approved are to be submitted to RTS. Research data collected, observation/field records, or results/reports will not become part of any student, teacher, principal, or school record. All research data and records must be locked in a secure location.

The researcher will please submit quarterly activity status by email to Gail Flanagan, RTS starting November 1, 2004 until research activities at NPS are completed.

The school district recognizes the importance of effective, research-based instructional strategies, and supports faculty and staff in pursuit of advanced university degree and dissertation research. We look forward to your findings and contribution to improved instructional practice, program services, and student achievement.

Sincerely,
Gail Flanagan
Senior Coordinator, Research & Evaluations
Department of Research, Testing & Statistics, Norfolk Public Schools
Office: 757-628-3852 Fax: 757-628-3925 gflanaga@nps.k12.va.us

Appendix L

Institutional Review Board Approval Letter

Virginia Tech

VIRGINIA POLYTECHNIC INSTITUTE
AND STATE UNIVERSITY

Institutional Review Board

Dr. David M. Moore
IRB (Human Subjects)Chair Assistant Vice President for Research Compliance
CVM Phase II- Duckpond Dr, Blacksburg, VA 24061-0442
Office: 540/231-4991; FAX: 540/231-6033
email: moored@vt.edu

DATE: September 24, 2004

MEMORANDUM

TO: Travis W. Twiford Educational Leadership & Policy St. oc 0302
Mark Tavernier

FROM: David Moore

SUBJECT: **IRB Exempt Approval:** "Formative Reading Program Assessment: An Interim Tool
For improvement" IRB # 04-453

I have reviewed your request to the IRB for exemption for the above referenced project. I concur that the research falls within the exempt status. Approval is granted effective as of September 24, 2004.

cc: File