STATEWIDE EARLY-GRADE READING ASSESSMENT PRACTICES

IN THE UNITED STATES

By

CHRISTINA LYNNE GILCHRIST

A dissertation submitted in partial fulfillment of
the requirements for the degree of

DOCTOR OF PHILOSOPHY IN EDUCATION

WASHINGTON STATE UNIVERSITY
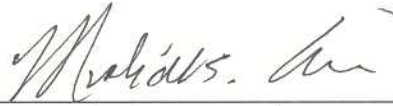Department of Educational Leadership and Counseling Psychology

AUGUST 2004
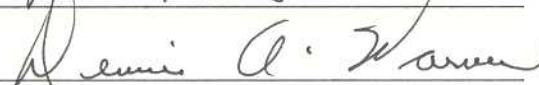
To the Faculty of Washington State University:

The members of the Committee appointed to examine the dissertation of Christina

Lynne Gilchrist find it satisfactory and recommend that it be accepted.

_____
Chair

_____

_____

ACKNOWLEDGEMENT

STATEWIDE EARLY-GRADE READING ASSESSMENT PRACTICES

IN THE UNITED STATES

Abstract

by Christina Lynne Gilchrist, Ph.D.
Washington State University
August 2004

Chair: Michael S. Trevisan

With the passing of the No Child Left Behind Act in 2001, more and more states are
implementing statewide assessments. Such assessments typically include a reading component.
Critics claim that reading assessments are incompatible with contemporary reading theory
(Levande, 1993; Winograd, Paris, & Bridge, 1991), and that the results of such standardized tests
do not provide useful guidance for instruction.

This study investigated statewide early-grade reading assessment practices in the United
States. A survey of assessment practices was developed according to Dillman's (2001) Tailored
Design Methods prescriptions. Individuals who indicated that they had knowledge of early-
grade reading assessment practices were selected from each state department of education,

Twenty-seven of the fifty-states, plus the District of Columbia, responded. Among
responding states, most passed legislation requiring statewide assessment with early-grade
reading assessment as a component of that program. States work with testing publishers to
develop assessments that align with the state standards. A couple states assess reading skills
such as phonemic awareness, letter recognition, writing, listening, prior knowledge, and
vocabulary, which correlate with a more contemporary theory of reading.

Most of the assessments are mandated. In all states that responded, third graders are given the entire state test between October and April. More states now also mandate assessing K-2. One state described its assessment as ongoing. Three states administer only a portion of the assessment.

These state assessments are standardized—administered by teacher or other district personnel to groups or individuals. The assessments are primarily scored by an outside agency, often the publisher that created the assessment. Available scores include the percent meeting the standard, the raw score, the percentile rank, the standard score, and other scores. State representatives, along with the outside agencies, then develop reports and disseminated them via parent teacher conferences and press releases.

The stakes are high—the federal legislation requires tough sanctions for students, teachers, and school districts if they perform poorly on the new assessments.

Modest progress has been made toward improving the fit between reading assessment practice and theory. However, making assessment results more useful to teachers and students would ultimately increase the assessments' utility and cost-effectiveness.

TABLE OF CONTENTS

LIST OF TABLES

Dedication

I dedicate this work to my husband and daughter—my two sweeties. I also dedicate this to my

Father in Heaven. Too many miracles came my way, all of which were beyond my own abilities.

May my work be acceptable before Him.

CHAPTER ONE

INTRODUCTION

Overview

This dissertation is a report of a study of early-grade reading assessment practices in the United States. Factors that might influence the conceptualization of this study include societal, intellectual, professional, political, and research considerations. This chapter provides a background and justification for studying early-grade reading assessment in light of these considerations.

Societal Considerations

Reading has become increasingly important since the turn of the century. Jennings (1969) presents an overview of society's norms and the circumstances that enable reading pedagogy and development in the United States. During colonial days, religious books were used to teach doctrine and reading skills at the same time. Printing presses allowed the number of books, as well as other types of publications, to be accessible for the common person. Reading became a skill that anyone could acquire. As the workplace became more technologically sophisticated, reading became necessary for employment. Despite the need for adequate reading skills, about 38% of students read poorly (Christie, 2002). Poor reading skills lead to failure in other academic courses (Rosier, 2002). In addition, those who fall behind in reading rarely catch up (Torgensen, 1998).

Another societal influence on early-grade reading testing is that testing itself has become an integral part of our society. According to Wixson, Valencia, and Lipson (1994), the U.S.

spends $700 million and 20 million school days testing. Although testing has always played a role in education, there seems to be an increased number of mandated tests. Logically, there has been an overall trend to assess reading statewide, using standardized assessments. Value placed on the assessment results, in general, seems to be indifferent to heavy criticism against it—creating an inelastic demand. In fact, the amount of testing has actually increased, regardless of educational domain. As schools are obliged to teach students to read, there must be some proof that the schools are meeting that obligation. At the intersection of these two developments lies early-grade reading assessment—the focus of this study.

Professional Considerations

Professional developments within the educational realm make early-grade reading assessment practice a controversial topic. In particular, two major professional matters affect the practice—our concept of reading and the nature of the reading assessments themselves.

First, developments in our concept of reading have changed. Levande (1993) agrees with the appraisal that reading is dynamic and interactive. Standardization makes it difficult to assess such dynamics. Others claim that these tests do not measure "real reading" (Curtis & Glaser, 1983; Glaser, 1998; Vanneman & White, 1998). According to Carroll (1979) critics claim that reading tests should measure how one "knows a language." McAuliffe (1993) characterizes reading as idiosyncratic and questions the value of standardized test results. Tierney (1998) further clarifies the problem: ". . .what may serve as a vehicle for uncovering literacies of one student may not be a satisfactory method for uncovering those of another student. . ." (p. 382). Such views of reading are categorized as holistic.

Second, some researchers claim that standardized reading tests do not reflect our understanding of cognitive development. Henk and Rickelman (1992) investigated whether states were moving toward holistic statewide reading assessments—which would reflect the contemporary understanding of reading. At that time, some states were reported to be interested in implementing assessments based on contemporary research findings and theories. However, the authors concluded that reading assessment has not kept pace with instruction (see also Hodges, 1997 and Tierney, 1998). Thus, changes to our understanding of reading. To date, these concerns have not been addressed.

Political Considerations

Recent political trends at various levels underscore the importance of studying early-grade reading assessment. Policymakers have increasingly relied on single test scores to make decisions, leading to high stakes situations for students, teachers, and schools (Haertel, 1999; International Reading Association, 1999). The amount of trust placed in assessment scores is never questioned (Haertel, 1999; Hodges, 1997). Furthermore, the external users of assessment results have received greater priority than internal users (Wolf, LeMahieu, & Eresh, 1992). According to Jongsma and Farr (1993), the increase in assessment activity has pitted assessment users against each other. Traditional assessments do not meet the needs of teachers (Guthrie & Mitchell, 1994), and classroom assessments do not provide the sense of accuracy and authority that policymakers desire (Tierney, 1998). There seems to be no common ground.

Not only do policymakers' interests lie in assessment results, but also they are becoming very involved with the reading assessment process (Orlich, 2000). Wolf, LeMahieu, and Eresh (1992) point out that various professional organizations cooperated in developing state standards

and assessments. The federal government has joined in, by passing the No Child Left Behind Act of 2001. This act requires states to implement standards based on research, and assess whether student achievement meets those standards (Clarke & Gregory, 2003; United States Congress, 2001). This legislation gives research a new role and, therefore, value in policymaking (Wixson & Pearson, 1998). With more and more testing resulting from the federal government's call to implement and assess new standards statewide, a snapshot of current practice could present a view of the status of current practice.

Significance of the Study

The trend in state level reading assessment affects many different groups at different levels. Most importantly, the use of assessment results affects students, who are being tested more often at younger ages. Decisions based on results from such tests include promotion or retention, entrance into an educational institution, (International Reading Association, 1999), placement (Tyson & Coker, 1991), and whether or not a student receives a certificate of mastery (Riddle Buly & Valencia, 2002). Darling-Hammond (1994) noted that although the assessment type would change, uses of the results would remain the same.

Reading assessment practices also impacts teachers and their practice. The International Reading Association (1999) cites that teachers' salaries in the future may depend on performance on the assessments; furthermore, personnel decisions may also make based on assessment data. The current assessment trend also impacts instruction. According to Grindler and Stratton (1992) the implementation of high-stakes assessments has led to drill and practice exercises to improving students' performance on those assessments.

The value of assessment results depends on getting the right information to the right people at the right time for the right purpose. At the national level, results are analyzed as group comparison data to administrators long after the assessment for the purpose of sorting, screening, and placing students. In contrast, at the classroom level, rich information is available to teachers immediately, which then is used to guide instruction as any errors occur (Wixson, Valencia, & Lipson, 1994). Thus, the two ends of the spectrum are broad indicators for accountability and academic discourse (Dressman, 1999). Haertel (1999) questions whether external assessment information can inform instruction at all.

Knowledge and understanding about current reading assessment practices across the nation can help policymakers and practitioners explore reasons why assessment information may not be as useful as expected.

Purpose

The main purpose of this study is to investigate the practices of each state regarding early-grade reading assessments. Essential components of the practice consist of policymaking mechanisms and assessment methods.

Overview of the Methodology

This study was primarily descriptive in nature. Data were collected via survey research, which inquired into statewide assessment practices.

Delimitations

There are a few limitations imposed on this study. First, the scope is limited to Kindergarten through 3rd grade initiatives, which assess early-grade or "beginning" reading skills. Second, the study focuses on state-level policies and practices. Third, the study focuses on practices—not scores or standards.

Definitions of Key Terms

Certain terminology is used throughout the literature that requires some clarification. *Early-grade reading* refers to pre-kindergarten through the third grade, which can be characterized as basic reading skills and perception; later reading is done for comprehension and other higher-level tasks (Chall, 1967). Also, *assessment* refers to the "specification and verification of problems for the purpose of making different kinds of decisions" (Salvia & Ysseldyke, 1998, p. 22). *Large-scale standardized reading assessment*, then, refers to state or national assessments on reading that are administered to all participants in the same conditions and manner (Wheeler, & Haertel, 1993).

Report of the Study

This study is divided into five chapters. Chapter One contains the background and justification for the research. Chapter Two is comprised of the literature review. Chapter Three provides an overview of the Tailored Design Method of survey research and describes how it was implemented in the course of this study. Chapter Four summarizes the results of the survey. Chapter Five summarizes the findings and suggests implications and future directions for research of early-grade reading assessment practice.

CHAPTER TWO

LITERATURE REVIEW

Overview

Much of early-grade reading assessment practice relates to other aspects of education. This literature review situates the study within the contexts of educational, assessment, and reading assessment practices in the United States. This chapter contains three sections. First, a historical background of reform is presented. Second, a synthesis of theoretical and empirical research is provided. Finally, a critique of the literature is presented.

Historical Background

Education Reform

Early American education was characterized by small communities or groups hiring a teacher for the purpose of reinforcing principles of Christianity and preparing future leaders (Urban & Wagoner, 1996). Roles of various stakeholders were kept fairly separate. Teachers were deemed responsible for teaching, but the onus for learning lay on parents and students (Peters, Wixson, Valencia, & Pearson, 1993). The federal government remained relatively uninvolved (Bourque, 1999); local control prevailed (Kirst, 1995).

The progressive movement of the late nineteenth century, considered the first wave of reform, had a social and an economic focus (Berube, 1994). The purpose of schooling, according to Atkinson (2002), was to alleviate social problems brought on by the arrival of many immigrants. Increasing enrollment and dissatisfaction with administration and curriculum led to some of the changes (Urban & Wagoner, 1996). The federal government mostly collected data;

states were still primarily responsible to provide public education and make most decisions. They passed laws compelling minors to attend school (Urban & Wagoner, 1996). More and more politicians became involved in the education realm (Kirst, 1995).

Societal changes stemming from World War I also affected educational policies. Intelligence testing became prevalent. The structure of the current educational system was formally instituted (Urban & Wagoner, 1996).

Later, the Great Depression certainly had some effect on schools. According to Urban and Wagoner (1996), some systems shut down completely. Many educators joined teacher's organizations, such as the AFT and the NEA. Most attempts to reform American schools at this time failed (Urban & Wagoner, 1996).

Then, during the 1950's, the Soviet Union and the launching of Sputnik presented a perceived threat to the United States (Roeber, 1999). This threat became the grounds for passing the National Defense Education Act of 1965, which focused on math and science education to bolster national defense (Bourque, 1999).

The second wave of education reform, in the 1960's and 1970's, called for equal educational opportunity, beginning with Brown v. Board of Education of 1954 (Urban & Wagoner, 1996). Congress passed the National Defense Education Act in 1958, or the "NDEA" to fix differences in achievement. Thus Congress, along with federal agencies, exerted power to help social problems via legislation and regulation (Roeber, 1999). This resulted in the new notion that teaching must account for outcomes (Peters, et al., 1993), as opposed to actions of parents or students. The most influential legislation was the Elementary and Secondary Education Act of 1965. The titles attempted to prevent poverty from affecting access to education (Fuhrman, 2001). Districts were held accountable (Clune, 2001).

Berube (1994) calls the 1980's the beginning of the third major wave of education reform—the standards reform movement. The 1983 publication, *A Nation at Risk*, sparked many states' standards reforms. According to Clune (2001), the theory behind standards-based reform was this:

> *Standards-based reform* (SR), through its purposeful activities, leads to *Standards-based policy* (SP), which leads to . . . *rigorous, implemented Standards-based curriculum (SC)* for all students, leading to Measured *high student achievement* (SA) in the curriculum as taught. (p. 15)

The assumption was that, for the U.S. to stay economically competitive, it would have to adopt policy-level curriculum approaches similar to that of our international competitors. Furthermore, policymakers assumed that such high stakes would naturally lead to better instruction and, therefore, better performance (Riddle Buly & Valencia, 2002). Both states and districts initiated systemic changes (Fuhrman, 2001). There was a shift to hold schools accountable for student learning (Clune, 2001).

More recent attempts to hold schools accountable are seen in the reenactment of the ESEA and the No Child Left Behind Act. Both legislative acts, implemented at the national level, concern setting national standards and mandate assessing whether students are meeting those standards (Chudowsky & Pellegrino 2003).

Assessment reform

A discussion of assessment reform obviously begins with tracing the origins of assessment itself. Then, a brief history of assessment in the United States will be followed by descriptions of assessment types and purposes.

Assessment is a fundamental component of education. According to Walberg, Haertel, and Gerlach-Downie (1994), since assessment purposes tie directly to every level of decision-making, assessment reform closely follows education reform.

Janesick (2001) provides a succinct chronology of assessment in the United States. In the late 1800's, Binet became noted as the father of intelligence testing in France. In the U.S., Thorndike developed standard scales, beginning the norm-referenced trend—the first wave of reform. Soon, tests were developed for all school subjects. Meanwhile, Horace Mann and Samuel Gridley Howe convinced the Boston School Committee to administer an impartial written test to ascertain whether students were taught—a focus on assessing the quality of instruction (Rothman, 2001). The aim was to hold teachers accountable for learning.

The early 1900's became known as the scientific era. During this time, the United States employed the Army Alpha (Jones, 1993; Urban & Wagoner, 1996), which was a standardized intelligence test. In the military, the Alpha test scores were used to ascertain who would become an officer or match recruits with jobs (Janesick, 2001). Adoption of Thorndike's opinion that "whatever exists, exists in a quantity" (Rothman, 2001, p. 422), was likewise reflected in increased testing. This was the beginning of large-scale group testing, focusing on testing intellectual skills. With large enrollments, efficiency became the primary criterion. Hardaway (1995) noted that the sorting of students through the use of such assessments, in light of compulsory education, led not only to classification, but also to segregation.

By the 1920's, the Scholastic Aptitude Test (SAT) was developed, with no particular curriculum in mind. The use of such standardized intelligence tests became more popular. Soon, they had almost the sole source of assessment information (Janesick, 2001; Ravitch, 1995). The underlying assumption was that intelligence was innate and fixed. Therefore measuring intelligence would predict achievement.

As Sputnik and ESEA affected the education reform movement, the government required assessment data to reassure the country of positive changes. Evaluation became synonymous with standardized achievement test scores (Perrone, 1977).

The second wave of assessments, according to Janesick (2001), occurred during the 1960's and 1970's. Standardized tests were still used, but the reasoning behind testing became more purposeful. These assessments more directly related to curricula and purported to determine whether or not educational objectives were met (Rothman, 2001). At this time, Gardner's theory on multiple intelligences sparked a flurry of new assessments. These assessments moved away from the notion of fixed intelligence as a predictor of achievement. But in holding states accountable, the indicators were mainly input-based (Goertz, 2001). States expanded their assessment systems by developing the Iowa Test of Basic Skills and the California Achievement Test (Roeber, 1999). One major criticism was that the states did not link them to the curriculum (Jennings, 1998). The tests were usually some form of minimum competency assessment (Rothman, 2001).

In the third wave, primarily in response to *A Nation a Risk,* politicians called for an increase in testing to measure achievement of more rigorous standards (Janesick, 2001; Rothman, 2001). Some states took the lead in this "excellence" reform. Stemming from the larger education movement itself, assessments were developed to investigate whether students

met the new standards. In particular, statewide assessment has flourished, since the overarching goal of the national policy was to ensure that states set those standards (Tucker & Codding, 2001). The onus of learning was place on the shoulders of the teachers and the schools. Media reports and sanctions for the schools reflected this change (Goertz, 2001). Along with the development of standards came the rise of alternative and/or "authentic assessments." The purpose of these new assessments was to provide a "real-world, complex" assessment (Walberg, Haertel, & Gerlach-Downie, 1994).

Trends toward more accountability eventually led to the No Child Left Behind Act, which mandates annual testing between grades 3 and 8 by the year 2005. Currently, the legislation requires sanctions for poorly performing schools include state takeover or the withholding of federal funds (Solomon, 2002).

Reading assessment reform

Reading assessment reform does not have three stages, as education and assessment reform movements do. Rather, it has two identifiable movements. An overview of the origins of reading assessment is followed by a description of the two reading assessment reform movements.

The practice of standardized reading assessment also dates back to the early 1900's, when Thorndike used standard scales to measure reading and mathematics (Janesick, 2001). Reading and mathematics were the only subjects assessed, besides intelligence, at that time. Reading tests focused on decoding, sight vocabulary, and comprehension of isolated sentences or short paragraphs (Winograd, Paris, & Bridge, 1991). According to Peters et al. (1993), until the 1960's, much of the reading assessment practice remained largely untouched since the 1910's.

Pearson and Dunning (1985) characterized the tests as "objective, machine-scorable, multiple choice tests that exhibited high reliability" (p. 305).

The criterion-referenced movement, beginning in the 1960's, is considered the first wave of reading assessment reform. At that time, the movement was influenced by the mastery learning movement in psychology. Federally mandated programs, such as Reading is Fundamental and Head Start (Peters et al., 1993), were accompanied by the development of new assessments to evaluate those programs. Another criterion-referenced, non-standardized instrument was the Informal Reading Inventory (IRI), which was in instrument that initiated the informal assessment frenzy (Warncke & Shipman, 1984).

The issuance of a report titled "Becoming a Nation of Readers: The Report of the Commission on Reading" by Anderson, Hiebert, Scott, and Wilkinson (1985), along with an improved model of reading led to the construction of the National Assessment of Educational Progress (Peters et al., 1993).

In response to the political impetus, Michigan and Illinois initiated statewide reform, which included reading assessment. Construction and development of these assessments departed from the old model of reading. The new assessments included factors such as prior knowledge or reading strategies (Henk & Rickelman, 1992). Furthermore, they aligned more closely with the curricula (Peters et al., 1993).

Most recently, the No Child Left Behind Act of 2001 has required states to implement standards for reading and assessing whether or not students meet those standards (U.S. Congress, 2004). This new mandate compels states to comply by implementing statewide assessment of reading.

Theoretical Literature

Various models underlie some of the current policies and practices of testing. Only a few models merit attention here, since they speak specifically to reading assessment practice.

The first model of literacy assessment by Winograd, Paris, and Bridge (1991) states "children must become strategic readers" (p. 112). Such strategy requires skill, motivation, and independence. The premise is that, in order to enable student to improve strategy, they must monitor (or assess) their own development. A teacher's sense of the student's status and progress would be documented informally. Any documentation would enhance or even replace the typical reporting formats of test scores and report cards. Thus, assessment, if used properly, would enable students to become strategic, and therefore "good" readers.

A second model, specified in Wixson, Valencia, and Lipson (1994), purports that tensions among reading assessment uses and users could cease if internal assessment is embedded within external assessment. In this model, the focus of the utility is greater for the teacher than for the policymaker, although such data is still useful to the policymaker. The decisions would be made through coordinated efforts of policymakers and teachers. The underlying theory is that, since internal and external purposes need not be mutually exclusive, efforts to negotiate purpose, design, and content should result in improved curriculum and instruction. This should then result in improved reading and learning, as reflected by assessment results.

Both models proclaim teachers and students to be the ultimate consumers of the assessment results. They also prescribe ways that results of classroom assessments could be more attractive to policy-makers. Whether or not politicians would take note and allow results from classroom assessments to guide their decision-making is an entirely separate issue.

One major weakness of these models is that they assume data from internal/informal assessments would satisfy external purposes. If the internal and external purposes are at such odds, then perhaps they also need rather different data. This may or may not preclude using the same assessment results. Another weakness is that the models also ignore the fact that proximal data, such as student observations, could hardly be compiled statewide with any ease. Such an endeavor could amount to a statewide portfolio. Application of these models would require serious consideration about how to compile such data. No such model exists to date.

Empirical Literature

Most of the reading assessment literature was prescriptive in nature, consisting of opinions regarding educational issues. Very little, however, could be categorized as truly empirical (i.e. experimental) research; rather, they are descriptive. Furthermore, the design is typically either a case study or a national survey.

First, Hodges studied the trust that school personnel invested in standardized and alternative assessments (1997). Findings indicated that the well-known achievement tests were trusted, possibly due to familiarity. Taken to its logical conclusion, then, non-standardized assessments would not be as trustworthy as scores from those standbys. Another study, spearheaded by Atkinson (2002) examined the relationship between the politically charged nature of the statewide assessment and the relationship between policy-level administrators and local implementers. The author found, through case study, that the logical divide between the two levels of users created an antagonistic relationship preventing useful and meaningful progress in developing an assessment program.

Linn and Haug (2002) researched the stability of accountability scores over time. The authors discovered that the estimates resulting from measures could be unstable. Class or school size significantly impacted stability of scores.

Another study on assessment scores, by Riddle Buly and Valencia (2002), described how reporting simple percentage scores concealed information vital to understanding the reading difficulties. The authors combined multiple measures of reading's key components, including word identification, phonemic awareness, comprehension, reading fluency, and vocabulary. Then they conducted both a factor analysis, to account for variance on the WASL, and a cluster analysis, to ascertain whether the pattern fit the majority of students. The findings indicated that a more sophisticated analysis of the multiple scores yielded information that could effectively guide instruction. The researchers mentioned that the WASL was comparable to the NAEP, but did not provide any further evidence of its comparability. Furthermore, the sample was limited to students scoring below proficiency. It is not clear whether such analysis would prove useful for the majority of students.

Allington and McGill-Franzen (1992) conducted a case study that examined reporting practices of seven school districts on a statewide assessment. These districts excluded various low-scoring students when reporting percentages of students by placing them in special education or retaining them in the same grade. The authors found that only a third of the original cohort took the exam. Although the analysis is simple, its ramifications are enormous.

One crucial piece—Henk and Rickelman (1992)—investigated whether or not the content that the states' newly mandated assessments reflected the more holistic model of reading. The authors found that states were beginning to use performance assessments in some areas. They still used scores from national level standardized tests to assess reading, however. Yet, the

authors failed to find significant indication that states implemented assessments based on the new definition of reading. Henk and Rickelman argue, "it may be too soon to expect pervasive changes to occur in statewide reading tests" (p. 79).

Summary and critique

Much of the literature consisted of case studies, which described implementation of individual states. Although many of the case studies provided detail, it did not give a clear picture of assessment practice from top to bottom. The resulting collection of results does not amount to a true cross-sectional picture of assessment practice.

One study, conducted by Henk and Rickelman (1992), which aimed to provide a cross-sectional view, surveyed assessment practices in the U.S. The authors did not find appreciable progress toward implementing assessments aligned with current reading models. Perhaps, as the authors argued, the cause truly did lie in bad timing—the authors may have prematurely researched the implementation. Or, the states truly did not begin adopting the new assessments.

One consistent finding, however, was that the needs of the most important users—teachers and students—were either ignored or overridden. Furthermore, that, even if their needs were considered, the assessment results did not adequately satisfy them.

Conclusion

It has been twelve years since Henk and Rickelman (1992) surveyed statewide reading assessment practices. As noted in the other literature, various users were not satisfied with the systems in place. With further impetus of recent legislative acts at both state and federal levels, it seems appropriate to re-examine the practice of statewide reading assessment practices.

There are many topics in reading literature that might prove useful when examining states' practices regarding early-grade reading assessment. From these issues stem eight research questions:

1. Do states have early-grade reading assessment practices?

2. If, so, what was their rationale?

3. How did they choose an assessment?

4. What were the assessment's characteristics?

5. How were the assessments administered?

6. How were the assessments scored and interpreted?

7. How were the assessment results used?

8. What were the consequences for performance on the assessment?

This study aims to answer these research questions, as described in the chapter on methodology that follows.

CHAPTER THREE

METHODOLOGY

Overview

This chapter is divided into three sections. Section one contains a description of the participants in the study. Section two gives an overview of social exchange theory. The final section presents the sequence and timing of survey implementation.

Synopsis of Social Exchange Theory

Data collection for this study was guided by the "Tailored Design Method" (Dillman, 2000). This method embodies social exchange theory—a perspective that informs the question construction, survey design, and implementation phases of the research.

The Tailored Design Method applies social exchange theory to respondents' behavior in survey research. According to Dillman (2000), "…*rewards* are what one expects to gain from a particular activity, *costs* are what one gives up or spends to obtain the rewards, and *trust* is the expectation that in the long run the rewards of doing something will outweigh the costs" (p. 14).

Application of these concepts to respondent behavior logically leads to the need for increasing rewards. Standard courtesies, such as showing positive regard and saying thank you, earn first mention. Other social notions such as asking the respondent for advice, appealing to values shared by various groups, and giving social validation, are all appropriate things to communicate in contacts with the respondent.

Ideas for decreasing costs for respondents include avoidance of subordinating language, embarrassment, and inconvenience. This means that the surveys should be respectful, appear short and easy, and have a self-addressed stamped return envelope.

Means of establishing respondents' trust could involve inserting a token of appreciation or securing sponsorship for the research by a legitimate authority. Also, printing personalized cover letters on letterhead stationery conveys an interest in that respondent's reply by a trusted authority.

Dillman (2000) uses these premises to inform question and survey construction as well as data collection. Highlights of his strategies follow.

Principles of Question Construction

Principles of excellent question construction include proper word choice, minimizing bias, accurate measurement, and good technical construction. Sound measurement stems from appropriate application of these principles. Each of the following examples illustrates their utility.

First, the researcher must choose few, but simple and precise words. Clear language will allow the respondent to accurately understand the question. Also, the researcher must avoid using either too vague or too specific quantifiers, which will hinder accurate and quick responses.

The author proposes that bias can be avoided by using various tactics. This includes balancing positive and negative categories, allowing for both neutral and undecided responses, and using equal comparisons within a scale and mutually exclusive categories.

Furthermore, the researcher can minimize mis-measurement by removing double negatives. A research should not ask respondents to make unnecessary calculations. Also, removing double-barreled questions helps respondents identify the key issue at hand.

Application of these principles increases the overall quality of questions individually. It is also important to achieve good measurement in the survey as a whole.

Principles of Survey Construction

Three steps are prescribed for survey construction, which aim to provide a common stimulus for each respondent. First is defining a navigational path—including navigational symbols, use of page space, and placement of crucial details that enable the respondent to stay on task. One suggestion is to place instructions exactly where the information is needed, rather than in a separate booklet.

Second, consistent font size and color can help respondents distinguish between questions and answers. Also, appropriate headings can help respondents to keep focused on the topic and correctly interpret questions within that section.

Third, major visual changes direct respondents to change their navigational patterns. The best example is skip patterns. Otherwise, respondents may continue on to questions that would not be appropriate or necessary for them to answer.

Finally, multiple reviews and revisions of the resulting document provide ample opportunity to fix any problems. Pre-testing by colleagues can increase assurance that questions are understood and that the right questions are asked. A final proofing by someone who has not been involved can prevent mistakes.

Even a well-designed survey, however, could still be rendered useless if the right people do not receive and return the survey. Therefore, once a survey is developed, due care must be given to implementing data collection.

Survey Construction

The survey consisted of six parts: assessment choice, assessment characteristics, assessment administration, scoring and interpretation, use of assessment results, consequences.

The first section investigated whether or not states were required to implement statewide early-grade reading assessment, or if they planned to do so in the future. The purpose was to get a sense of what states used to implement statewide reading assessment, the general policy mechanisms, date of the requirement, and which stakeholders were involved in making the decision to implement.

The second section focused on the characteristics of the assessment(s) adopted by the state for early-grade reading assessment. Basic information such as the assessment name, author, publisher, publication year, and number of forms was requested. Some questions covered the assessment's technical qualities, such as reliability and validity measures. Other questions indicated assessment techniques and content. The third section attended to issues of administration. Questions centered on whether the assessment was administered to groups or individuals, how many times in a school year students were assessed, whether the whole assessment was administered or not, which administration modes were employed, who administered the assessments, and how long the assessment lasted.

The fourth section on scoring and interpretation of the assessment information clarified who scored the assessment, what types of scores were available and how they were summarized,

and who developed the report. The fifth section concerned how assessment results were used. Information regarding how the results were communicated, who was expected to use the results and whether those users were both trained and informed of the results were requested. The sixth section indicated whether or not consequences existed for students, teachers, and districts, based on assessment results.

Four versions of the survey were created: a self-administered paper survey, an MS Word electronic form, a PDF copy, which was available on the web, and a telephone interview protocol (see Appendix A). Only consent formats and administration modes changed among the versions. Paper copies of both cover letters and surveys were developed and sent by regular mail using department letterhead and stationery. The body of e-mails contained the same language as the paper cover letters with a MS Word copy of the survey attached. It also referenced the URL where an identical PDF version of the survey could be downloaded. The 15-20 minute telephone interview protocol began with a verbal consent statement, followed by the same questions as in the other survey formats.

As prescribed, the survey was pre-tested. A local school teacher in the area with expertise in K-3 reading was asked to provide feedback on answer options prior to data collection. An individual not involved in the survey development was asked to review and critique the survey.

Participants

State departments of education were the intended unit of analysis, since they are the agencies authorized by the state to enact legislative and administrative mandates. Surveys were sent to each State Department of Education and the District of Columbia, totaling 51 participants.

As each state department was organized differently, no single title or position uniformly served as an appropriate contact. The researcher selected potential respondents based on their expertise or experience regarding statewide early-grade reading assessment practice.

Sequence and Timing

Research activities occurred during the 2003-2004 school year. Each data collection stage closely followed Tailored Design Method prescriptions (Dillman, 2000), which emphasizes multiple, personalized contacts with participants. Therefore, a series of contacts were planned— an initial recruiting phone call, pre-notice letter, initial survey, follow-up postcard, e-mail follow-up and replacement surveys, and special telephone contacts.

The researcher compiled a list of potential contacts of results from a web search with the keywords "state" "department" and "education." Phone calls and e-mails were sent to verify the suitability of each person as an information source. The researcher sent pre-notice letters to verify contact information. A few weeks later, the researcher mailed the first wave of paper surveys. Next, postcards, gently reminded respondents to complete and return the survey, or thanked those who had already responded. Then, when very few responses were received, special steps were taken to persuade non-respondents to reply. E-mail requests for either a confirmation of the respondent's ability to provide information on the topic, or for a referral to a more appropriate person accompanied an electronic copy of the survey. Finally, phone calls permitted the researcher to collect the data over the phone.

CHAPTER FOUR

RESULTS

Overview

Twenty-seven states responded for a 54% response rate. Three declined to participate entirely—two due to time constraints. The remaining states did not respond to any stimulus.

Data Analysis

Responses to individual questions, with related questions whenever possible, were transformed into tables and matrices. To identify trends among state practices, simple descriptive central tendency measures were calculated. The results are presented by survey category and question.

Assessment Choice

Question 1: Does your state require statewide assessment of early-grade (or kindergarten through 3rd grade) reading? Yes or No? If there are no current plans for early-grade reading testing, do plans exist for early-grade reading testing in the future?

Twenty-five states supplied information regarding their own assessment practices. Of these, 14 states (56%) currently require statewide assessment of early-grade reading. Two states (8%) did not have current assessments, but indicated plans in the future to assess early-grade

reading (see Table 1). Nine states had neither current statewide early-grade reading assessments

nor future plans to assess (36%).

Table 1

Requirements for Statewide Early-Grade Reading Assessment

State	Current[a]	Future[b]
AK	0	0
CO	X	N/A
CT	0	0
DE	X	N/A
FL	X	N/A
GA	X	N/A
ID	X	N/A
IL	0	0
MD	0	X
ME	0	0
MI	X	N/A
MO	0	0
MS	X	N/A
MT	0	0
ND	0	X
NE	0	0
NH	X	N/A

Table 1 (continued)

States Requiring Current or Future Reading Assessment

State	Current[a]	Future[b]
NY	0	0
OR	X	N/A
SC	0	0
TN	X	N/A
TX	X	N/A
WA	X	N/A
WI	X	N/A
WV	X	N/A

Note. X = Yes, 0 = No, N/A = Not Applicable; n = 25.

[a] States that currently require statewide assessment of early-grade reading. [b] States which are planning to require statewide assessment of early-grade reading.

Question 2: Which of the following issues was most important in deciding to implement

statewide assessments of early-grade reading?

For the 14 states that reported having statewide assessment of early-grade reading, respondents ranked issues important to deciding whether to implement the statewide reading assessment (Table 2). Accountability ranked highest among responding states. General state indicators, such as summarizing and predicting reading achievement, were ranked higher than screening and placement concerns.

Table 2

Ranks of Issues Important to Deciding to Assess Reading (n = 14)

Administration Mode	Mode[a]	Minimum	Maximum
Summarize Achievement	4	1	5
Predict Achievement	3	0[b]	5
Refer to Services	1	1	5
Screen for Eligibility	1	0[b]	5
Accountability	5	1	5

Note. [a]These data were reverse-scored in order to preserve the intent of the question. That is, 5 = most important issue and 1 = least important issue. [b]0=no answer provided.

Question 3: Which of the following policy mechanisms were used to implement the early grade reading testing?

Legislative regulations led 13 of the 14 responding states to implement statewide reading assessment for (Table 3). Administrative regulations were used to implement assessments in 64% of the states and information dissemination led to implementation for 57% of the responding states. Finally, in only 14% of states that responded, financial incentives were identified as a policy mechanism used to encourage implementation. Another state mentioned that creation of some other structure was a mechanism used to implement early-grade reading assessment. No state reported the use of setting priorities for supporting activities as a policy mechanism.

Table 3

Policy Mechanisms Used to Implement Assessments by State

State	Information Dissemination	Financial Incentives	Administrative Regulations	Legislative Regulations	Creation of Structure	Setting Priorities
CO	X		X	X	X	
DE	X		X	X		
FL				X		
GA	X		X	X		
ID		X		X		
MI	X	X	X			
MS	X		X	X		
NH				X		
OR	X			X		
TN			X	X		
TX	X		X	X		
WA				X		
WI	X		X	X		
WV			X	X	X	
%	57%	14%	64%	93%	14%	0%

Note. X = Yes, Blank = No; n = 14.

Question 4: When did the official state requirement begin?

For states responding to this question, the months and years of actual or intended implementation of a statewide early-grade reading assessment ranged from January 1979 through September 2004 (See Table 4). The median start date was November 1998.

Table 4

Month and Year of Implementation of a Statewide Early-Grade Reading Assessment Program

State	Month and Year of Implementation
CO	2/1998
DE	9/2004
FL	3/2000
GA	7/2004
ID	9/1999
MI	9/1999
MS	7/2000
NH	5/1994
OR	3/1998
TN	3/1989
TX	1/1979
WA	3/1997
WI	9/1988
WV	9/2002

Note. n = 14.

Question 5: How involved were each of the following groups in deciding which assessment

should be used statewide?

Various groups are involved in making the decision to assess early-grade reading

statewide (see Table 5). These include state officials, school districts, parents, and teachers.

Both state officials and teacher groups were reported to be very involved among most states that

responded. Likewise, state legislatures were highly involved. School districts were somewhat

involved and parents were minimally involved, according to respondents.

Table 5

Level of Group Involvement in Deciding to Assess Reading Statewide

Group	Mode	Minimum	Maximum
State Officials	3	2	3
School Districts	2	0	3
Teachers	3	0	3
Parents	2	0	2
Other	2	0	3

Note. 3 = Very Involved, 2 = Somewhat Involved, 1 = Minimally Involved, 0 = Not Involved,

-1 = Don't Know; n = 14.

Question 6: Why was this assessment chosen?

Six of twelve states indicated that an assessment was created to align to state needs (See Table 6). Three of those states were required by legislation to create their own assessments, one of which employed an RFP (Request for Proposal) process. One state respondent mentioned that the assessment was developed, not chosen. Two states chose an assessment based on observation, survey, best practice or research. One state decided to create an assessment to guide instruction, which was a practice prior to the No Child Left Behind Act. Finally, one state declared "It was not a 'test' for young children. [The assessment] allows for students to demonstrate skills over time."

Table 6

Reasons Why Specific Assessments Were Chosen by State

State	Reason
CO	State legislation provided an option to build test on State Model Content Standards. It was decided to build a test to be aligned with the existing 4th grade reading assessment.
DE	It was not a "test" for young children. It allows for students to demonstrate skills over time.
FL	Not chosen—developed.
ID	The state assessment was based on the National Research Council's work *Preventing Reading Difficulties in Young Children*.
MI	Based on literacy, best practice, observation, survey.
MS	RFP Process

Table 6 (continued)

Reasons Why Specific Assessments Were Chosen by State

State	Reason
NH	It is a custom test that is designed to match the NH English Language Arts Curriculum Framework.
TN	This assessment was closely aligned to our state curriculum, and it was part of an assessment series that would be administered in grades 3-8.
TX	State law required that the Texas Education Agency develop its own tests; we don't require schools to administer commercially developed tests.
WA	There is a menu of assessments that a school can chose from.
WI	We did not choose an assessment. We created our own. The law requires that we develop & annually administer the assessment. Existing commercial tests were not adequate for our needs & WI has a belief in including WI educators throughout the process.
WV	The assessment program was implemented to guide student instruction, which was done prior to the No Child Left Behind Act.

Note. n = 14.

Question 7: Which curricular/instructional reading philosophy does your state promote?

Thirteen states provided information regarding which curricular/instructional philosophy they promote. Five states promote the Phonics/Whole Language reading or instructional philosophy. Two others promote Balanced Literacy. Two more promote the "5 Components" of reading, of which one is a local control state. One state specified the "Other" philosophy.

Another state mentioned that it was a local control state. One did not provide a description for an "Other curricular/instructional reading philosophy." Finally, one respondent stated that no philosophy was adopted by the state (Table 7).

Table 7

Percentage of States Adopting a Curricular/Instruction Philosophy

Philosophy	Percent
Phonics	0.0%
Whole Language	0.0%
Phonics/Whole Language	38.5%
Other[a]	54.0%
No Philosophy	7.5%

Note. n = 13. [a]Clarification provided for the "Other" category included Reading First, Balanced Literacy, and the 5 Components.

Question 8: How much influence did the state reading philosophy have on choosing the
　　　　assessment?

Finally, states reported the level of influence that their chosen curricular/instructional philosophy had on choosing the assessment. Six of the 12 responding states (50%) indicated that the philosophy had a lot of influence in the choice of assessment. Two mentioned that their philosophy had some influence (16.7%). Only one said that the philosophy had very little influence on assessment choice (Table 8). Three stated that the level of influence was not applicable (25%). One state did not respond to this question.

Table 8

Level of Influence of Reading Philosophy on Assessment Choice

Level	Percent
A lot of Influence	50.0%
Some Influence	16.7%
Very Little Influence	8.3%
No influence	0.0%
Don't Know	0.0%
Not Applicable	25.0%

Note. n = 12.

Assessment Characteristics

The first portion of the Assessment Characteristics asked for unique information identifying the assessment. Table 9 summarizes responses to the following questions:

Question 9: What is the full name(s) of the assessment(s) used?

Question 10: Who is the author of the assessment?

Question 11: What is the name of the publisher?

Question 12: What year was the assessment published?

Question 13: How many forms of the assessment are available?

Question 15: Is a technical report or manual available concerning how the assessment was

developed? If yes, please describe.

Some states were required by state law to create their own assessments. Of those (see Table 8), seven asked commercial publishers to create assessments (50.0%, n = 14), one by the Request for Proposals (RFP) process. Three states wrote their own assessments (21.4%). Two states had a menu of commercial assessments to choose from (14.2%).

Table 9

Assessment Names, Authors, Publishers, and Dates, Number of Forms, and Manuals by State

State	Name	Author	Publisher	Year	Forms	Manual
CO	CO Student Assessment Program	CTB/McGraw Hill	CTB/McGraw Hill	1998	2	X
DE	Work Sampling Assessment System	Samuel Meisels	Pearson Early Learning	2002	1	
FL	Florida Comprehensive Assessment Test	Harcourt Educational Measurement	Harcourt Educational Measurement	1998	1	X
GA	Georgia Knowledge Assessment Program - Revised	Georgia State University	J. Popham	1990	1	X

Table 9 (continued)

Assessment Names, Authors, Publishers, and Dates, Number of Forms, and Manuals by State

State	Name	Author	Publisher	Year	Forms	Manual
GA	Criterion-Reference Competency Test (1-3)	Riverside Publishing	Riverside Publishing	2002	2	
ID	Idaho Reading Indicator	Waterford Early Learning/ID State Dept Ed	Waterford	2000	1	
MI	MI Literacy Progress Profile	MI Early Literacy Committee	MI Dept of Ed	1998	3	X
MI	Dynamic Indicators of Basic Early Literacy Skills (DIBELS)	Good, Simmins, & Kame'enui	University of Oregon	2002		
MS	MS Curriculum Test	CTB/McGraw Hill	CTB/McGraw Hill	2000	3	X

Table 9 (continued)

Assessment Names, Authors, Publishers, and Dates, Number of Forms, and Manuals by State

State	Name	Author	Publisher	Year	Forms	Manual
NH	NH Educational Improvement Assessment - End of 3rd Grade LA	Measured Progress Test Item Development	Measured Progress Test Item Development	Each Year	8	X
OR	Read/Literacy Knowledge & Skills Assessment	OR Dept of Ed	OR Dept of Ed	2003	20	X
TN	TN Comprehensive Assessment Program	CTB/McGraw Hill	CTB/McGraw Hill	2003	6	X
TX	Texas Assess of Knowledge and Skills	TX Ed Agency		Each Year	56	X
WA	Developmental Reading Assessment	Joetta Beaver	Pearson Early Learning	?		
	Sunshine Reading Assessment	Wright Group/McGraw Hill	Wright Group/McGraw Hill	?		

Table 9 (continued)

Assessment Names, Authors, Publishers, and Dates, Number of Forms, and Manuals by State

State	Name	Author	Publisher	Year	Forms	Manual
WA	Analytical Reading Inventory - 6th Ed.	Mary Lynn Woods & Alden J. Moe*	Prentice Hall*	1999*		
	Multilevel Academic Skills Inventory	Ken Howell	Psychological Corporation*	1989*		
	Qualitative Reading Inventory - 3rd Ed.	Lauren Leslie & Joann Caldwell	Allyn & Bacon	2001		
	Reading and Writing Assessment Folder	Dominie Press	Dominie Press			
WI	WI Reading Comprehension Test	WI Dept Pub Inst	WI Dept Pub Inst	2004	1	X
WV	WV Informal Reading Assessment	Beverley Kingry	WV Dept of Ed	2003	1	X

Note. X = Yes; Blank = No information provided; n = 14. *Some states did not disclose this information, which was found on Buros Institute of Mental Measurements Test Reviews Online at http://buros.unl.edu/buros/jsp/search.jsp.

Question 14: How would you describe the type of assessment?

Ten of the thirteen responding states used criterion-referenced assessments (76.9%).

Two described the assessment(s) as being both criterion and norm-referenced (15.4%). One state

described the assessment type as "Other" and explained "Teacher observation and evaluation"

(7.7%).

Question 16: Which of the following reliability measures are available? If an actual measure is

available for your test, please write the coefficient in the line provided (i.e. r =

+0.63).

Forty percent of responding states indicated use of the test-retest reliability, as showing in

Table 10 (n = 10). Fifty-six percent used internal consistency measures. Most states did not

provide actual reliability coefficients for assessments (60%). The KR-20, Cronbach's Alpha,

and Classical Item were specifically described. Coefficients, if levels were provided, ranged

from .86 through .91.

Table 10

Reliability Measure Types by State

State	Test-Retest	Parallel-Forms	Split-Half	Internal Consistency	Coefficient
CO	X				Cronbach's Alpha = 0.90
FL				X	.91
GA				X	KR-20= 0.85 and 0.94
ID	X				r=0.86
OR				X	Classical Item =0.92
TN				X	KR-20
TX				X	0.89
WA[a]					
WI	X			X	
WV	X				

Note. X = Yes; Blank = No; n = 10. [a]WA State has multiple assessments.

Question 17: Which of the following validity measures are available?

For validity measures of the assessments, some states provided specific types, while others mentioned simply that validity measures were available (See Table 11). Most states with an assessment also employed content validity measures (57%, n = 8). Three responding states (37.5%) used face validity measures, with one state using peer review while another correlated

current with later scores. Another three responding states addressed content validity, one using correlation with later scores, one with peer review, and another aligning with state content objectives. Another 37.5% determined predictive validity, two states correlating earlier scores with later scores. Three states measured concurrent validity, one of which used factor analysis. Two states also established construct validity (29%).

Table 11

Validity Measure Types by State

State	Face Validity	Content	Predictive	Concurrent	Construct
CO		Correlation with later scores			
FL				.84	.63
GA	Peer Review	Peer Review	Correlation with later scores		Factor Analysis
ID				Measured against ITBS Spearman's rho = .64	
OR	X	X			
TN	X	X	X	X	

Table 11 (continued)

Validity Measure Types by State

State	Face Validity	Content	Predictive	Concurrent	Construct
TX		X	At exit level only (High School)		At High School Exit
WV		Align 1:1 w/state content objectives			

Note. An "X" denotes validity was measured, but no description was given; n = 8.

Question 18: The assessment requires which of the following assessment techniques?

Fourteen states reported types of assessment techniques used. Table 12 shows that the majority (64%) of states used Pen-and-Pencil assessments. Forty-three percent used student products to assess reading; 36% using observation, 14% anecdotal records, and 7% chose "other" assessment techniques. No descriptions were available regarding those "other" techniques.

Table 12

Assessment Techniques Employed by State

State	Observation	Student Product	Pen-and Pencil	Anecdotal Records	Other
CO			X		
DE	X	X			
FL			X		
GA	X	X	X	X	
ID		X			
MI	X	X			
MS			X		
NH			X		
OR			X		X
TN			X		
TX			X		
WA	X	X		X	
WI			X		
WV	X	X	X		
%	36%	43%	64%	14%	7%

Note. X = Yes; Blank = No; n = 14. No descriptions for "Other" assessment techniques.

Question 19: Which of the following reading skills or behaviors are assessed?

All 14 states that responded assessed comprehension skills (Table 13). Most measured word recognition skills (71%). Many (53%) assessed oral reading skills while or evaluated word

attack skills and 43% assessed comprehension strategies. Fewer (31%) observed reading rates.

phonemic awareness, letter recognition, writing, listening, viewing, prior knowledge, and

vocabulary. The fewest states that responded to this question indicated that reading rate was

assessed.

Table 13

Reading Skills and Behaviors Assessed by State

c	Oral[a]	Skills[b]	Strat.[c]	Attack[d]	Recog.[e]	Rate[f]	Other[g]
CO		X			X		
DE	X	X	X	X	X		
FL		X	X		X		
GA	X	X		X	X		
ID	X	X		X	X	X	Phonemic awareness, & letter recognition
MI	X	X	X	X	X	X	
MS		X			X		
NH		X					Writing, listening, viewing
OR		X					
TN	X	X	X	X	X		
TX		X	X	X			

Table 13

Reading Skills and Behaviors Assessed by State

State	Oral[a]	Skills[b]	Strat.[c]	Attack[d]	Recog.[e]	Rate[f]	Other
WA	X	X			X	X	
WI		X	X				Prior knowledge
WV	X	X		X	X	X	Phonemic awareness, vocabulary
%	50%	100%	43%	50%	71%	29%	29%

Note. X = Yes, Blank = No; n = 14. [a]Oral Reading. [b]Comprehension Skills. [c]Comprehension Strategies. [d]Word-Attack Skills. [e]Word Recognition Skills. [f]Reading Rate.

Question 20: Are alternate forms of the assessment available? If yes, please describe.

Six state respondents indicated that an alternate form of the assessment was available (see Table 14). The alternate forms, including Braille and large print, plain language, or for progress monitoring, were typically limited to a very small percentage of the population,.

Table 14

Alternate forms available for the assessment

State	Alternate Form	Description
CO	X	Alternate for 2% population
DE		
FL		
GA		
ID		
MI	X	Progress monitoring
MS	X	Large Print/Braille
OR	X	Plain language available Extended Reading Assess
TN	X	Braille/Large Print
TX	X	For students whose ARD Committee decide TAKS is inappropriate
WA		
WI		
WV		5 levels mastery; after 3 attempts if not mastery
Total	42%	

Note. n=14.

Administration of the Assessment

Table 15 summarizes responses to the following questions:

Question 21: At what grouping level are students assessed?

Question 22: How many times is the assessment administered during the year?

Question 23: Which month(s) of the year is the assessment administered?

Question 25: Which part(s) of the assessment is/are used?

Question 27: How many minutes does the assessment last?

Seven of the 14 responding states administered reading assessments to groups (50%), while the other half administered assessments to individuals.

Table 15

Assessment Administration Details by State

State	Grouping	Months	Portion	Description	Minutes
CO	Group	February	Entire		120
DE	Individual	October, January, April	Part(s)	English/LA/Math/Personal/Social Development	N/A
FL	Individual	March	Entire		50
GA	Individual	March, April	Entire	Ongoing (Kindergarten)	120
ID	Individual	September, January, April 15th - May 15th	Entire		10
MI	Individual	September, January, May	Entire	Dibels 10, MLPP	depends
MS	Group	May	Entire		Un-timed
NH	Group	May	Entire		205
OR	Group	April	Entire	Optional on-line October-May	90

Table 15 (continued)

Assessment Administration Details by State

State	Grouping	Months	Portion	Description	Minutes
TN	Group	March Through April	Entire		185
TX	Group	March, April, July	Entire		120
	Individual	October, January,	Part(s)	All DIBELS, MASI-R	Various
WA		May		Fall, 1 min Jan, 1 min	
				Spring	
WI	Group	March	Entire		Un-timed
	Individual	At least once, up to 3	Part(s)	5 areas: phonemic	2 Overall
		times		awareness, phonics,	Estimated
WV				vocabulary, fluency,	
				comprehension	

Note. N = 14.

 September, January, March, April, and May are popular months to administer the assessments. Two states had administration during optional months. Two states required that students be assessed once, one of which allowed up to three optional re-assessments, with no preference for month. One state described assessment as ongoing.

 Most states administer the entire assessment (78.6%, n=14). Three states (21.4%) administer only a portion of the assessment. Parts of this assessment used include English, Language Arts, Math, Personal and Social Development, and the 5 areas (phonemic awareness, phonics, vocabulary, fluency, comprehension).

Three states had un-timed assessments. Another state estimated approximately 2 hours of assessment overall. The average duration for the assessment(s) is 108.9 minutes (n = 9, median = 120 minutes, mode = 120 minutes). One did not disclose timing.

Question 24: Under which conditions are kids assessed?

Thirteen state respondents indicated what conditions led to assessing students. Six responding states had mandatory assessments for each grade from kindergarten through the third grade. Another state required assessment during the second and third grades only. Five states mandated assessment during the third grade only. All thirteen states assessed during the third grade.

Table 16

Conditions Under Which Students Assessed

State	Kindergarten	1st Grade	2nd Grade	3rd Grade
CO				Mandatory
DE	Mandatory	Mandatory	Mandatory	Mandatory
FL				Mandatory
GA	Mandatory	Mandatory	Mandatory	Mandatory
ID	Mandatory	Mandatory	Mandatory	Mandatory
MI	Mandatory	Mandatory	Mandatory	Mandatory
MS			Mandatory	Mandatory
OR				Mandatory
TN	OTHER:	OTHER:	OTHER:	Mandatory

Table 16 (continued)

Conditions Under Which Students Assessed

State	Kindergarten	1st Grade	2nd Grade	3rd Grade
TX				Mandatory
WA	Mandatory	Mandatory	Mandatory	Mandatory
WI				Mandatory
WV	Mandatory	Mandatory	Mandatory	Mandatory
%	53%	53%	62%	100%

Note. N = 13.

Question 26: Which mode(s) of administration is/are used?

Assessment administration modes also varied by state (See Table 17). In 71% of the states that assess reading in the early grades, an assessment administrator read standardized instructions to groups. Fifty-seven percent read standardized instructions to individuals. Some states (43%) used self-administered assessments. Another 14% used undisclosed "other" assessment modes. Finally, one state used computer-screen administration (7%).

Table 17

Assessment Administration Modes by State

State	Self-Administered	Standardized to Group	Standardized to Individual	Computer Screen	Other	Description
CO		X				
DE					X	N/A
FL	X	X	X			
GA	X	X	X			
ID			X			
MI			X			
MS		X				
NH		X			X	Accommoda-tions are specified and allowed
OR		X	X	X		
TN	X	X				
TX	X	X	X			
WA			X			
WI	X	X				
WV	X	X	X			
%	43%	71%	57%	7%	14%	

Note. X = Yes, Blank = No; n = 14.

Question 28: Who administers the assessment?

Teachers administered the assessments in most states (92%, n = 14). Table 18 shows that district personnel administer assessments in 7 states (54%), while paraprofessionals administer in 4 states (31%). Other assessment administrators included certified administrators, assessment teams in the building, and principals (23%).

Table 18

Groups that Administer Assessments by State

State	Teachers	Paraprofessionals	District Personnel	Other
CO	X			
DE	X			
FL	X	X	X	
GA	X			
ID			X	
MI	X		X	
MS	X		X	Certified Administrators
NH	X			
OR	X	X	X	
TN	X			

Table 18 (continued)

Groups that Administer Assessments by State

State	Teachers	Paraprofessionals	District Personnel	Other
TX	X	X	X	
WA	X	X	X	Assessment Teams in the Building
WI	X	X	X	
WV	X			Principals
%	93%	36%	57%	21%

Note. X = Yes, Blank = No; n = 14.

Scoring and Interpreting Assessment Information

Question 29: Who, or what agency, scores the assessment?

Most states (42.9%) had some other agency or group score the assessments (Table 19).

These included CTB Scoring Center in Indianapolis, Harcourt Educational Measurement,

CTB/McGraw Hill, Measured Progress, Educational Data Systems, and Pearson Educational

Measurement. Some 14.3% of the state departments of education scored the assessments, while

another 14.3% had both the outside agency and the state department of education sore

assessments. Some states (14.3%) had teachers score the assessments. About 7.1% had schools

score; districts scored for another 7.1% of states.

Table 19

Scoring Agency or Groups by State

State	Agency or Group	Description
CO	CTB Scoring Center Indianapolis	Other
DE	Harcourt Educational Measurement	Other
FL	Pearson Educational Measurement	Other
GA	Georgia Department of Education	DOE
ID	District Personnel	District
MI	Teacher, data entered on website	Teacher
MS	CTB/McGraw Hill	Other
NH	Measured Progress	Other
OR	Educational Data Systems	Other
TN	State Department of Education and CTB/McGraw Hill	Other and DOE
TX	Texas Educational Agency and Pearson Educational Measurement	Other and DOE
WA	Individual Schools	Schools
WI	WI Dept of Pub Inst	DOE
WV	Teacher or Principal	Teacher

Note. n = 14.

Question 30: What types of scores are available?

Types of scores available differed by state (see Table 20). The majority of states presented the percentage of students meeting the standard (69.2%, n = 13). Many states (64.5%) displayed raw scores; another 53.8% used "other" types of scores, but did not specify the type. States calculated percentile ranks (46.2%). Standard scores were determined for 30.8% of the states and stanines in 23.1%. Only 1 state reported using the grade equivalent (7.7%).

Table 20

Scores Types Calculated

Type	Percent of States
Raw Score	64.5%
Percentile Rank	46.2%
Stanines	23.1%
Standard	30.8%
Grade Equivalent	7.7%
% Meeting Standard	69.2%
Other	53.8%

Note. n = 13.

Question 31: How are scores summarized?

Scores were aggregated at various levels. Table 21 shows that most states offered scores or state levels; 46.2% aggregated by classroom and 38.5% by some "other" level.

Table 21

Score Aggregation Levels

Type	Percent of States
Individual	84.6%
Classroom	46.2%
District	76.9%
State	76.9%
Other[a]	38.5%

Note. n = 13. [a]Other aggregation levels included schools and campus; one state reported having the districts decide aggregation levels.

Question 32: Who develops the report?

Different groups within states developed the assessment reports (Table 22). Most states had state representatives take part in report development (77% n = 14). Over half had an outside agency assist (54%). Fifteen percent of states had either teachers participate; another 15% of states had school district representatives work on the assessment report. One state's school administrators assisted in report development.

Table 22

Groups Involved with Assessment Report Development by State

Group	Percent
Outside Agency	57%
State Representatives	71%
School District Representatives	14%
School Administrators	7%
Psychologists	0%
Counselors	0%
Teachers	14%

Note. n = 13.

Assessment Results

Question 33: How are results communicated?

States communicated results of the statewide assessment differently (Table 23). Most states held parent-teacher conferences (92%, n = 14). Many states used press releases to share scores (77%). Thirty-eight percent of responding states held special meetings (38%). States who indicated "Other" dissemination types described the formats. One example is a Parent Report that is distributed; another state provides building, district, and state reports for Reading First schools. One state sends reports to superintendents and principals first and then the reports are given to the district for dissemination. Another state reported that information dissemination varied by school district.

Table 23

Method of Disseminating Assessment Results by State

Method	Percent of States
Parent-Teacher Conferences	92.3%
Press Releases	76.9%
Special Meetings	38.5%
Other	38.5%

Note. n = 13.

Question 34: Who is expected to use the assessment results?

States expected different groups to use the assessment results (see Table 24). Both school administrators and teachers were expected to use assessment results in all of the responding states (n =14). Ninety-two percent of the responding states expected state officials to use them; another 92% expected school districts to use results. Most states expected parents to use assessment results (85%). Many states (69%) expected counselors to use assessment results. Students were expected to use results in 54% of responding states, as well as psychologists in another 54% of states.

Table 24

Groups Expected to Use Assessment Results by State

Group	Percent of States
State Officials	92.3%
School Districts	92.3%
School Administrators	100.0%
Counselors	69.2%
Psychologists	61.5%
Teachers	100.0%
Parents	84.6%
Students	61.5%
Other	0.0%

Note. X = Yes, Blank = No; n = 13.

Question 35: How much training did each of the following groups receive regarding the

appropriate interpretation and use of the assessment results?

Thirteen states described various levels of assessment use training to different user

groups (see Table 25). States reported thorough training levels for state officials, school

districts, and school administrators. Counselors, psychologists, and teachers received some

training in assessment result use in responding states. Parents, students and "other" groups were

reported to have no training in using assessment results in almost all responding states;

Wisconsin, however, provided some training to state legislatures, state government, educational

organizations, and interest groups.

Table 25

How Much Assessment Use Training Groups Received by State

Group	Mode	Minimum	Maximum
State Officials	3	1	3
School District	3	2	3
School Administrators	3	1	3
Counselors	2	0	3
Psychologists	2	0	3
Teachers	2	0	3
Parents	0	0	3
Students	0	0	2
Other	0	0	2

Note. 3 = Thorough Training, 2 = Some Training, 1 = Minimal Training, 0 = No Training; n = 14.

Question 36: How informed are each of the following groups about the assessment results?

Furthermore, states inform various groups at different levels (Table 26). States report that state officials, school officials, and school administrators are all very informed of the assessment results (n = 14). Counselors, psychologists, teachers, parents, and students are somewhat informed about assessment results. No level of inclusion is given for "other" assessment result user groups.

Table 26

How Informed States Keep Groups of Assessment Results

Group	Mode	Minimum	Maximum
State Officials	3	1	3
School District	3	2	3
School Administrators	3	2	3
Counselors	2	0	3
Psychologists	2	0	3
Teachers	2	2	3
Parents	2	1	3
Students	2	0	3
Other	1	1	1

Note. 3 = Very Informed, 2 = Somewhat Informed, 1 = Barely Informed, 0 = Not Informed; n = 14.

Consequences

Table 27 summarizes the responses to questions 37 through 39:

Question 37: Are there consequences for STUDENTS based on their own performance on the reading assessment?

Question 38: Are there consequences for TEACHERS based on students' performance on the reading assessment?

Question 39: Are there consequences for SCHOOL DISTRICTS based on students' performances on the reading assessment?

Consequences imposed by states were based primarily on NCLB or by other state laws for students, teachers, and school districts. A few were positive rewards, but most states reported negative sanctions for poor performance.

Table 27

Consequences for Groups Based on Performance by State

State	Students	Teachers	School Districts
CO			District accreditation if results are not improving
DE	Individual Improvement Plans Developed		
FL	Possible Retention	District decision	District decision
GA	K-3 Promotion/Retention	NCLB/state law provide rewards/sanctions	NCLB/state law provides rewards or sanctions
ID			
MI	Referral for additional support	Varies by district	If funded through Reading First
MS	3rd grade benchmark assess for promotion		AYP; Priority Schools
NH			No, because of NCLB may be identified as school in need of improvement

Table 27 (continued)

Consequences for Groups Based on Performance by State

State	Students	Teachers	School Districts
OR			AYP implications; Report Card
TN		Teacher effect reports to determine "highly qualified" teacher status	If fail to meet AYP, placed on targeted or high priority status
TX	Possible retention in grade/remediation	Varies by dist, not at state level	If district is a consistently low performer, it may face state sanctions
WA		Interventions for students	NCLB Adequate Yearly Progress (AYP) status - as well as qualify for school improvement
WI	Considered for possible remediation if not minimal		If <80% Minimal, improvement plan

Note. n=14. Blank cells indicate no response.

Most states endorsed consequences for school districts, which were based on assessment performance (76.9%, n = 13). Such consequences were described as district accreditation or Adequate Yearly Progress (AYP) status implications, per NCLB or state law requirements. One state mentioned that there were implications for Reading First schools. Also, one state required an improvement plan if less than 80% met the minimum standard.

States mandated consequences for students depending on their assessment performance. Over half of states responding (53.8%) mentioned that individual improvement plans be developed, K-3 promotion or retention, referral for additional support, or a 3rd grade benchmark assessment for promotion serve as consequences for poor performance.

Five states (46.2%) described consequences for teachers for students' assessment results. Examples include rewards and sanctions for teachers as mandated by NCLB or state law. Other rewards or sanctions were decided at the school district level, such as teachers being identified as a highly qualified teacher, or required to determine interventions for students.

CHAPTER FIVE

DISCUSSION

Overview

This study investigated practices states used in implementing a statewide early-grade reading assessment. Findings are summarized by survey category. Next is a discussion of possible explanations and integration with past literature. Implications for the findings are offered. A brief discussion of limitations is followed by suggestions for future research.

Assessment Choice

Findings

Of the states that responded to the survey, 16 (59.3%) chose to assess early-grade reading statewide or were planning to implement a statewide assessment in the near future. Reasons for implementing focused on group-oriented goals of accountability and to summarizing student achievement. The remaining states had not implemented and indicated they had no plans to implement statewide early-grade reading assessment.

Policy mechanisms used to implement the assessment were primarily administrative and legislative administration and information dissemination, with implementation dates ranging from 1979 through 2004. Teachers and state officials were reported as highly involved in the process. States chose or created assessments that align with their state curriculum, which was highly influenced by their reading curricular/instructional philosophy.

Explanations and Integration with Literature

According to state respondents, the reason for implementing the statewide early-grade reading assessment rated on group-oriented issues, such as accountability and summary of student achievement, higher than individual issues, such as referral or screening. One explanation is that states wish to present their best face to the researcher, saying "Yes, we really are holding ourselves accountable." It is also possible states are primarily responding to political demands. That is, state politicians institute a new assessment program, regardless of its quality or value, as evidence accountability.

The tendency of responding states to assess early-grade reading is primarily reported to be in response to administrative and legislative regulations. An agency is expected to both comply with and create regulation; it theoretically does not act without authorization from existing legislation. These findings show an increase in implementation of state mandated assessments than previously reported (Henk & Rickelman, 1992; Linn & Haug, 2002), but were in line with the expectation of the International Reading Association (1999). However, these new assessments are focused on standards and criteria, rather than norm-referenced, as cited in Farr (1992).

A new trend seems to have emerged; states are creating assessments that are developed to align with the state curriculum and instructional philosophy. Maybe states, which are being held more responsible, are realizing that information from existing assessments had not informed their decisions well. That is, they may be developing ways to collect data that will enable them to make better decisions. Such circumstances are consistent with critics of the earlier accountability state mandated assessment movement, who claimed that the assessment results measured progress towards unknown ends.

Implications

In agreement with Farr (1992), this wave of new assessments is likely to continue. The tendency toward faddism and the political reality of election years have a significant impact on the adoption and implementation of early-grade reading assessment.

First, it is important to determine whether the current movement is just a fad. One assumption made by policymakers is that, if current statewide assessments do not meet the current information needs, a *new-and-improved* assessment will. Come election time, the policy is narrowed to who can provide the most *appealing*, rather than the most *appropriate*, means of collecting information that informs policy-level decisions. Such mandated policy may force assessments to serve external purposes may prevent teachers from having enough time to collect the assessment information they truly need.

Another political pressure is the short terms between election years—new assessments may not around long enough to adequately judge its utility and value. Politicians favoring one assessment have much at stake in upcoming election years and abandon an assessment before it can be evaluated.

One assumption is that if scores increase, then students are reading better. Madaus, as cited in Allington and McGill-Franzen (1992), argues that such inferences should not be made. Even if statewide assessments have been fully implemented, they may not be adequately catching students at risk. Practices of grade retention and exclusion would prevent poorly performing students from participating at all. Alternatively, if teachers' judgments are necessarily better, such assessments fail to serve as the safety net. The problem then becomes the back flips that schools and teachers feel they must take in order to meet the new *statistical* standard, rather than the *academic* standard.

Unfortunately, developing instruments that align with current theories regarding early-grade reading is only a relatively recent development. If assessment content has truly dictated instruction, then students would have been taught according to older models of reading. Furthermore, the logical conclusion would be to "improve" their reading by aiming to score better using that same model.

Another problem is that assessment development requires quite a bit of cooperation and negotiation. States need results that inform their decisions; teachers need information on which to base instructional decisions. Yet, since legislation and regulation require such high-stakes reporting, the results are communicated in statistics not useful to teachers, who need it the most. Perhaps the application or modification of reading assessment models (Allington and McGill-Franzen, 1992, Henk & Rickelman, 1992, Riddle Buly & Valencia, 2002, Wixson, Valencia, & Lipson, 1994) would meet the needs of both internal and external users might alleviate the tensions between the various users of assessment results.

As more and more states choose to implement statewide early-grade reading assessments, there is no need to reinvent the wheel. Instead, it is time for states to share the wealth, as well as the lessons learned.

Assessment Characteristics

Findings

Half the number of responding states chose to create a state assessment with the assistance of a commercial publisher; some wrote their own assessments; others used existing assessments. The majority of the assessments were criterion referenced—many aligning with their own state curriculum. Some states provided reliability measures with coefficients ranging

69

from 0.86 - 0.94. Validity measures used varied from peer review, alignment with content objectives, factor analysis, to correlation with other scores.

The assessments by far used pen-and-pencil assessment techniques. States used student products and observation techniques less frequently, and anecdotal records least often. Assessment domains always included comprehension. Word-attack and word recognition skills were assessed, as well as comprehension strategies. Oral reading was sometimes assessed, while reading rate was rarely assessed. A few states focused on phonemic awareness and letter recognition, prior knowledge, and vocabulary. There were alternate forms available for special needs populations, including Braille and large-print or plain language.

Explanations and Integration with Literature

Half of responding states used commercial publishers to assist in the development of their assessment. One reason that states may turn to these publishers is that they do not have the time or the resources to develop the assessment themselves. An alternative explanation may be that the states do not have the expertise to fully develop the assessment. The literature typically does not address the role of commercial publishers in statewide assessment development.

Another development is the shift toward using criterion-referenced assessments, rather than norm-referenced assessments. This is consistent with the trend noted by Linn and Haug (2002). However, the assessments are intended to align with the state curriculum *by design*. Perhaps the states are seeing the value of not only information about how students are compared to each other, but how they are meeting the standards. This practice is very similar to the aim of the earlier standards-based assessment movement (Henk & Rickelman, 1992), but with a twist. States may now be articulating the standards by defining the state curriculum.

70

When the coefficients for psychometric properties of assessments were actually provided, they were fairly high. One cannot determine why coefficients were not provided in the remaining cases. One explanation could be that the respondent did not know what they psychometric properties were. Or, perhaps the coefficients were lower and the respondent did not wish to provide such information. Alternatively, psychometric properties may not be valued information regarding early-grade reading. This corresponds with the argument that criterion-referenced assessments are incompatible with standard measures of concurrent validity.

Pen and pencil assessments were by far the most used method of assessing reading. Wixson, et al. (1994) noted that valuing efficiency and economy leads to the use of group, paper and pencil, short answer tests. This is a likely interpretation of these data, since statewide assessments must be administered to a vast number of students. They also include observation, which is supported by Doak and Chapman (1994).

States reported that domains included strategies and prior knowledge in some states. Such a practice reflects recent reading models more closely than found in previous literature (Henk & Rickelman, 1992).

Implications

The commitment to developing a "new" assessment may necessitate a departure from familiar quality indicators. This leaves a serious gap in the knowledge base regarding what would, or should, constitute evidence of validity. That is, it would not make sense to measure concurrent validity of a new reading assessment with that of an assessment based on a different model of reading.

One effect of adopting a ordaining a particular assessment is that it would standardize the results, making it easy to define performance. However, this would also mean that potentially useful data would not be valued by the state—only the statistics produced by the test publishers. Even if the assessment could yield valuable, rich data regarding student reading achievement, such information would be lost in highly aggregated reports.

The emerging trend of aligning assessments with a more current model of reading improves upon previous practice. This may lead to more useful information simply because it measures more appropriate reading variables. Again, it would only be useful if reports were developed to communicate such information.

Commercial publishers, who may have a significant influence in curriculum, also heavily influence what gets assessed. Policymakers must know the effect publishers have on the assessment purposes. Policymakers, then, must be HIGHLY conversant in assessment issues and understand the implications of involving commercial publishers.

Taking standardized pen and pencil assessments for reading may cause interpretation problems. That is, the skill of taking the assessment itself may confound the measurement of reading. A student who cannot read a test cannot take the test. Yet, a student who is accustomed and proficient at taking tests, but may not read well, could score high.

Finally, respondents did not provide detailed validity and reliability measures for these assessments. Without some evidence that the assessments provide trustworthy (valid and reliable) data. One cannot sensibly base good decisions on bad data. Perhaps it is simply too early to implement the assessment programs, because the technical quality needs to be refined. Or, perhaps states should not implement the poor assessments.

The bottom line is how do we really know when a child has been "left behind"? The two ends of the spectrum seem to be either laissez faire or high government involvement. The introduction of legislation and regulation necessarily forces accountability matters to be directed toward the state, which is not necessarily the most important stakeholder. As teachers spend more and more time assessing for external purposes, they are less able to make time to assess for instructional purposes that could make the most difference.

The implications for states' failure to report reliability and validity measures have two extremes. At one, the respondent may not have understood that actual coefficients were requested. Or, perhaps such information was not available to the respondent at the time he or she was filling out the survey. Such circumstances would not warrant alarm. However, on the other hand, states may have simply refused to provide a coefficient because they could have been lower than hoped. Or, worse yet, perhaps such figures did not even exist. If states focus on testing without evidence of the quality of the assessment, such information will be ultimately useless.

Administration

Findings

Most of the responding states administered the entire assessment once or twice a year for most states. For states administering the assessment only once, it was administered late in the school year (February, March, April, or May). Only 1 described assessment as "on-going." The average testing time was 108.9 minutes, while one was not timed.

Assessment was mandatory for *all* students in grades K-3 for just about half of the responding states. The others did not mandate the assessment until usually the 3rd grade. In the

73

administration, most often states reported that teachers read standardized instructions to a group, and sometimes read standardized instructions to an individual. Occasionally the assessments were self-administered. Sometimes school administrators and paraprofessionals were involved, including principals.

Explanations and Integration with Literature

State respondents are reporting that statewide assessment is scheduled throughout the year; some even report continuous assessment. This departs from the expectation of assessing only late in the year as mentioned in (Perry, 2000). Policy makers may see the value of smaller bits of information over time to note trends. Or perhaps they feel that more information is better information.

Responses indicate a trend toward assessing early-grade reading as early as kindergarten among responding states. These states may be trying to "catch" at risk-students "before they fall." Or they may feel pressured to assess earlier to show that they are compliant with NCLB regulation—a sort of an overachiever approach.

States are reportedly administering standardized assessments to groups and individuals; some are self-administered. This is no surprise, because of funding and staffing limitations, as mentioned in the literature. Only a few are using more diagnostic tools. Perhaps they are truly looking for more useful information, and wish to just monitor that teachers are administering and collecting that useful information for their own use.

Implications

The more involved the state is in setting standards and directing assessment decisions, the more training should come from the state. However, it is not known whether they truly show

their support with commensurate funding. Appropriate funding would allow proper

implementation and continuing support at all levels—state, district, school, classroom, and home.

If the assessments are administered too late in the year, it may be too late to do anything

about it (see also Grindler & Stratton, 1992). Teachers then would pass at-risk students on to the

next grade. If the only purpose of the assessment was to evaluate the statewide progress, that

would be fine. If it were to serve more diagnostic purposes, however, then reports late in the

year are useless.

Even if assessments information arrives at appropriate times, assessing kindergarteners

may be inappropriate. Standardized testing may confound cognitive development with actual

reading skills (Paris, Lawton, Turner & Roth, 1991). That is, one could potentially mistake a

students' familiarity with tests for a students' actual ability to read.

In short most any format has its shortfalls; giving standardized instructions to a group

may save time and money, but may result in information with poor diagnostic qualities.

However, to gain such depth and breadth of information would be nearly impossible. There is an

obvious cost-benefit analysis required for such decisions, which is frequently tilted toward

cutting costs. Such practice disregards the most important purpose—to make sure that no child

truly gets left behind. Following proper assessment practice at all levels might improve current

practice.

Scoring and Interpretation of Assessment Information

Findings

Other agencies scored the assessment most often, sometimes in conjunction with the state

departments of education. A variety of scores were calculated. The most frequently calculated

scores include raw score and the percent meeting the standard. Most often these scores were summarized at the state, district, and individual levels. State representatives and outside agencies wrote the reports. About half had teachers participate. Very few had school district representatives or school administrators assist with report development.

Results were communicated with parent-teacher conferences in many states. A high percentage also used press releases. Various reports were disseminated by district or building. Almost all major stakeholders were expected to use the assessment results. Psychologists and students were expected to use results by 61% of the responding states. State, district, and school administrators were trained in using assessment results. States reported that parents and students received no training. State, district, and school officials were reported to be very informed of the results. All other groups were somewhat informed.

Explanations and Integration with Literature

Assessments were described as being scored at an external location—scored elsewhere by agencies, sometimes with state departments of education. This is probably because publishers price scoring as part of a package deal. Furthermore, personnel have already spent too much time administering it, especially if it was to individuals.

Scores now include percent meeting the standard—as mentioned in earlier standardized assessment literature, but in now in more states. One reason may be that it is popular—showing that states adhere to the standards reform movement. Or perhaps the state does not need a higher level of measurement that nominal. That is, as long as the state can know how many kids are at risk in various districts, the states could then work with those districts directly to improve the situation.

State representatives were primarily in charge of reports. Makes sense, at the state level, it is mostly for state-level decisions. The representatives would know whether reports were intended for their own consumption or for their constituents.

Scores are aggregated at state, district, and individual levels. If districts are administering, certainly there will be an interest in district performance—especially with AYP status tied to results. Parents want to hear results of how their kid is doing. The state wants to keep a handle on things, since they are primarily responsible for making decisions at the state level.

Reporting at parent-teacher conferences and press releases were given as dissemination techniques. Things have not changed, and may be out of the hands of states. The press can be very demanding.

Parents and students were reported as receiving very little training. Very few teachers receive pre-service training in assessment in literature. Perhaps it is assumed that the numbers speak for themselves. Or, there may not be enough resources, including time, to train various assessment user groups. Another explanation may be that one test is administered just like any other, or scores are similar across any assessment.

Implications

Scores have little meaning without interpretation (MacGinitie, 1993). As long as the users involved know how to interpret them, scores would not be problematic. This begs the question of who, then, should interpret. The teacher has much more information about how Sally or Johnny reads on a regular basis. A system might allow snapshots to become state-level indicators to be combined with more extensive indicators. Such a system would require more

effort to summarize, but could make much more sense to all parties. Assuming the teachers implement the state curriculum in the classroom, it would be logical that there would be a high correlation between scores on the state assessments and classroom assessments.

Tying assessments to the state curriculum may assist state officials to make decisions about the curriculum. Furthermore, it relieves the problem of "narrowing the curriculum" because anyone teaching to the test would therefore be teaching the curriculum.

Consequences

Findings

Consequences for student performance on the assessment affected students, teachers, and school districts. Some states report that assessment results are used to determine whether students are promoted or retained. Other consequences include remediation, development of individual improvement plans, and referral for additional support. Teachers also face consequences for student performance. For the most part, the district decides the consequences for teachers. In one state, it helps determine the "highly qualified teacher" status. Finally, school districts are also affected by results. Consequences range from being placed on targeted or high priority status for failing to meet the No Child Left Behind Adequate Yearly Progress (AYP) standard, facing undisclosed state sanctions, losing district accreditation, or developing an improvement plan.

Explanations and Integration with Literature

In line with previous findings in the literature (Farr, 1992; International Reading Association, 1999), big decisions are still being based on single assessment scores, at all levels. But now, federal and state legislation and regulation are involved. Perhaps the main reason for implementing the assessment is to remain compliant. One state maintained that the assessment effort began prior to NCLB.

Implications

If the assessments were intended to truly give a state-level indicator, then individual reports would not be appropriate, even if individual results could be calculated. Out of sheer habit, reports may still be presented at the individual level, regardless of stated purposes and the level of sensitivity of the measures. Decisions made with report data aggregated for purposes other than intended (Tyson & Coker, 1991)—leads to high stakes. However, if the best use of results is to inform curriculum and instruction, then might consider using matrix sampling.

Limitations of the Study

Design and Internal Validity

First of all, it was extremely difficult finding "the one" who *should* respond to the survey. In a few states, two persons within the department argued between themselves about who should complete the survey. This might indicate a lack of leadership or vision in implementing statewide early-grade reading assessment. Or perhaps a completely different position within the department would provide more useful responses.

Another difficulty was finding "the one" who *would* answer. Non-response could mean that states were not interested in the topic. Or, perhaps the study was still premature. An

alternative explanation could be that respondents found the study threatening—should anyone find the truth about their practice, the states could be in trouble.

Additionally, policy implementation is quick. States could be taking more appropriate steps between the time respondents replied and the time this study is reported.

External Validity and Generalizability

The small number of respondents significantly limited the generalizability of results. A few problems resulted. First, it is difficult to say that there would was truly a trend, since only about half of the states responded. Second, the participants were specialist within state departments of education. These respondents may have a significantly different perspective than districts and teachers. The analysis of implementation would be limited then to how state departments of education perceive the assessment movement. Actual implementation decays as it moves down the line.

Analysis and Power

This survey relied on self-report mechanisms. It is possible that respondents provided socially desirable answers. The topic of early-grade reading assessment was a much more sensitive topic than originally anticipated. Some of the responses seemed to be bragging, such as "Based on literacy, best practice, observation, survey." Another state's "It was not a 'test' for young children. It allows for students to demonstrate skills over time" indicated the negative connotation of testing. However, the concepts of "testing" and "demonstrating skills over time" are not necessarily mutually exclusive. The difference, then, lies in the use of the results.

Furthermore, responses to the survey were aimed to provide a comprehensive view of the practice. This would prevent an in-depth analysis of the practices, which could allow a possible diagnosis of the problems.

Future Directions

Since state departments of education are state agencies, it is likely to have both legislative and regulative requirements. A better understanding of how the decisions are made would give a sense of who the assessment users are and what their needs would be. This would allow for greater analysis and insight into the political realities and means to address those realities.

It may also be useful to investigate statewide implementation from other assessment users—teachers and students. If they were considered equally valid users, then their insights would be just as valuable.

An obvious follow-up study is to see how results are actually used. In light of the harsh consequences of not meeting Adequate Yearly Progress in the No Child Left Behind arena, there may be some highly politicized and visible reporting of results.

Conclusion

In short, early-grade reading assessment in the US can be characterized as chaotic. States may be implementing assessments…but may not know what or how. Of the responding states, only a few seem to be developing assessments based on a more current model of reading.

Prior to implementation, states must attend to the needs of groups of assessment information users. They must also acknowledge that teachers and students may be the most

important users of the assessment information. Accountability to students, would then translate

into accountability at all levels.

REFERENCES

Allington, R. L., & McGill-Franzen, A. (1992). Does high stakes testing improve school effectiveness? *ERS Spectrum, 10*(2), 3-12.

Anderson, R. C., Hiebert, E. H., Scott, J. A., & Wilkinson, I. A. (1985). *Becoming a nation of readers: The report of the Commission on Reading.* Washington, DC: National Institute on Education.

Atkinson, T. S. (2002). "We're not just whistling Dixie": Policymakers' perspectives on state education reform. *Reading Research and Instruction, 41*(4), 289-308.

Berube, M. R. (1994). *American school reform: Progressive, equity, and excellence movements, 1883-1993.* Westport, CT: Praeger.

Bourque, M. L. (1999). The role of the National Assessment of Educational Progress (NAEP) in setting, reflecting, and linking national education policy to states' needs. In G. J. Cizek (Ed.), *Handbook of educational policy* (pp. 213-249). San Diego, CA: Academic Press.

Carroll, D. (1979). Standardized tests: What they can and cannot do. In M. Raggett, C. Tutt, & P. Raggett (Eds.), *Assessment and testing of reading: Problems and practices* (pp. 45-61). London: Ward Lock Educational.

Chall, J. S. (1967). *Learning to read: The great debate.* New York: McGraw-Hill.

Christie, K. (2003). Getting policymakers literate on literacy. *Phi Delta Kappan, 84*(1), 6-8.

Chudowsky, N., & Pellegrino, J. W. (2003). Large-scale assessments that support learning: What will it take? *Theory into Practice,* 75-83.

Clarke, M., & Gregory, K. (2003). This issue: The impact of high stakes testing. *Theory Into Practice, 42*(1), 2-3.

Clune, W. H. (2001). Toward a theory of standards-based reform: The case of nine NSF statewide systemic initiatives. In S. H. Fuhrman (Ed.), *From the capitol to the classroom:*

Standards-based reform in the states (pp. 13-38). Chicago: National Society for the Study of Education.

Curtis, M. E., & Glaser, R. (1983). Reading theory and the assessment of reading achievement. *Journal of Educational Measurement, 20*(2), 133-147.

Darling-Hammond, L. (1994). Performance-based assessment and educational equity. *Harvard Educational Review, 64*(1), 5-30.

Dillman, D. A. (2000). *Mail and internet surveys: The Tailored Design Method* (Second ed.) New York: John Wiley & Sons.

Doak, J. L., & Chapman, A. D. (1994). Educational reform in early elementary assessment. *Journal of Instructional Psychology, XXI*(1), 8-13.

Dressman, M. (1999). On the use and misuse of research evidence: Decoding two states' reading initiative. *Reading Research Quarterly, 31*(3), 258-285.

Farr, R. (1992). Putting it all together: Solving the reading assessment puzzle. *The Reading Teacher, 46*(1), 26-37.

Fuhrman, S. H. (2001). Introduction. In S. H. Fuhrman (Ed.), *From the capitol to the classroom: Standards-based reform in the states* (pp. 1-12). Chicago: National Society for the Study of Education.

Glaser, D. R. (1998). *Comprehensive literacy plan grades K-3 for the state of Idaho.* Boise, ID: Lee David Pesky Center for Learning Achievement.

Goertz, M. E. (2001). Standards-based accounability: Horse trade or horse whip? In S. H. Fuhrman (Ed.), *From the capitol to the classroom: Standards-based reform in the states* (pp. 39-59). Chicago: National Society for the Study of Education.

Grindler, M. C., & Stratton, B. D. (1992). Whole language assessment. *Reading Improvement, 29*(4), 262-264.

Guthrie, J. T., & Mitchell, A. (1994). Performance assessments in reading and language arts. P. Afflerbach & B. Kapinus (Eds.), *The Reading Teacher, 48*(3), 266-271.

Haertel, E. H. (1999). Performance assessment and education reform. *Phi Delta Kappan, 80*(9), 662-666.

Hardaway, R. M. (1995). *America goes to school: Law, reform, and crisis in public education.* Westport, CT: Praeger.

Henk, W. A., & Rickelman, R. J. (1992). Trends in statewide reading assesment: A closer look. *Reading Horizons, 33*(1), 67-82.

Hodges, C. A. (1997). How valid and useful are alternative assessments for decision-making in primary grade classrooms? *Reading Research and Instruction, 36*(2), 157-173.

International Reading Association. (1999). High-stakes assessments in reading. *The Reading Teacher, 53*(3), 257-264.

Janesick, V. J. (2001). *The assessment debate: A reference handbook.* Santa Barbara, CA: ABC-CLIOs.

Jennings, F. G. (1969). What is reading? In A. R. Bintner, J. J. Dlabal, Jr., & L. K. Kise (Eds.), *Readings on reading* (pp. 3-17). Scranton, PA: International Textbook Company.

Jennings, J. F. (1998). *Why national standards and tests? Politics and the quest for better schools.* Thousand Oaks, CA: Sage.

Jones, L. V. (1993). School achievement trends for black students. In B. R. Gifford (Ed.), *Policy perspectives on educational testing* (pp. 19-73). Boston: Kluwer Acadmic Publishers.

Jongsma, E., & Farr, R. (1993). A themed issue on literacy assessment. *Journal of Reading, 36*(7), 516-517.

Kirst, M. W. (1995). Who's in charge? Federal, state, and local control. In D. Ravitch & M. A. Vinovskis (Eds.), *Learning from the past: What history teaches us about reform* (pp. 25-56). Baltimore: Johns Hopkins University Press.

Levande, D. (1993). Standardized reading tests: Concerns, limitations, and alternatives. *Reading Improvement, 30*(2), 125-127.

Linn, R. L., & Haug, C. (2002). Stability of school-building accountability scores and gains. *Educational Evaluation and Policy Analysis, 24*(1), 29-36.

MacGinitie, W. H. (1993). Some limits of assessment. *Journal of Reading, 36*(7), 556-560.

McAuliffe, S. (1993). A study of differences between instructional practice and test preparation. *Journal of Reading, 36*(7), 524-530.

Orlich, D. C. (2000). Education reform and limits to student achievement. *Phi Delta Kappan, 81,* 468-472.

Paris, S., Lawton, T. A., Turner, J. C., & Roth, J. L. (1991). A developmental perspective on standardized achievement testing. *Educational Researcher, 20*(5), 12-20, 40.

Perrone, V. (1977). *The abuses of standardized testing.* Bloomington, IN: Phi Delta Kappa Educational Foundation.

Perry, A. (2000). Coaching for a reading test? Yes, it is possible and necessary. *Reading Improvement, 37*(3), 119-125.

Peters, C. W., Wixson, K. K., Valencia, S. W., & Pearson, P. D. (1993). Changing statewide reading assessment: A case study of Michigan and Illinois. In B. R. Gifford (Ed.), *Policy perspectives on educational testing* (pp. 295-391). Boston: Kluwer Academic Publishers.

Ravitch, D. (1995). The search for order and the rejection of conformity: Standards in American education. In D. Ravitch & M. A. Vinovskis (Eds.), *Learning from the past: What history teaches us about school reform* (pp. 167-190). Baltimore: Johns Hopkins University Press.

Riddle Buly, M., & Valencia, S. W. (2002). Below the bar: Profiles of students who fail state reading assessments. *Educational Evaluation and Policy Analysis, 24*(3), 219-239.

Roeber, E. D. (1999). Standards initiatives and American educational reform. In G. J. Cizek (Ed.), *Handbook of educational policy* (pp. 151-181). San Diego, CA: Academic Press.

Rosier, P. W. (2002). Our goal: 90 percent reading. *The School Administrator, January 2002,* 20-23.

Rothman, R. (2001). One hundred fifty years of testing. In L. Iura (Ed.), *The Jossey-Bass Reader on school reform* (pp. 419-433). San Francisco: Jossey-Bass.

Salvia, J., & Ysseldyke, J. (1998). *Assessment* (Seventh ed.) New York: Houghton Mifflin.

Solomon, P. G. (2002). *The assessment bridge: Positive ways to link testing to learning, standards, and curriculum improvement.* Thousand Oaks, CA: Corwin Press.

Tierney, R. J. (1998). Literacy assessment reform: Shifting beliefs, principled possibilities, and emerging practices. *The Reading Teacher, 51*(5), 374-390.

Torgensen, J. K. (1998). Catch them before they fall. *American Educator, 22*(1), 32-39.

Tucker, M. S., & Codding, J. B. (2001). Setting high standards for everyone. In L. Iura (Ed.), *The Jossey-Bass Reader on school reform* (pp. 470-482). San Francisco: Jossey-Bass.

Tyson, D., & Coker, D. R. (1991). Methods of assessment: Are they appropriate for young children? *Reading Improvement, 28*(2), 96-99.

United States Congress. (2001). No Child Left Behind Act of 2001. In *Public Law print of PL 107-110* (chap.) Retrieved July 14, 2004, from http://www.ed.gov/policy/elsec/leg/esea02/index.html

Urban, W. J., & Wagoner, J. L., Jr. (1996). *American education: A history.* New York: McGraw-Hill.

Vanneman, A., & White, S. (1998). Long-term trends in student reading performance. *NAEP Facts, 3*(1), 2-6.

Walberg, H. J., Haertel, G. D., & Gerlach-Downie, S. (1994). *Assessment reform: Challenges and opportunities.* Bloomington, IN: Phi Delta Kappa Educational Foundation.

Warncke, E. W., & Shipman, D. A. (1984). *Group Assessment in Reading: Classroom Teacher's Handbook.* Englewood Cliffs, NJ: Prentice-Hall, Inc.

Wheeler, P., & Haertel, G. (1993). *Resource handbook on performance assessment and measurement: A tool for students, practitioners, and policymakers.* Livermore, CA: EREAPA Associates.

Winograd, P., Paris, S., & Bridge, C. (1991). Improving the assessment of literacy. *The Reading Teacher, 45*(2), 108-116.

Wixson, K. K., & Pearson, P. D. (1998). Policy and assessment strategies to support literacy instrudction for a new century. *Peabody Journal of Education, 73*(3-4), 202-227.

Wixson, K. K., Valencia, S. W., & Lipson, M. Y. (1994). Issues in literacy assessment: Facing the realities of internal and external assessment. *Journal of Reading Behavior, 26*(3), 315-337.

Wolf, D. P., LeMahieu, P. G., & Eresh, J. (1992). Good measure: Assessment as a tool for educational reform. *Educational Leadership, 49*(8), 8-13.

APPENDIX A

COVER LETTERS

SURVEY PRENOTICE

States across the nation are battling poor reading achievement levels among their

students. In attempt to identify students who are at-risk for poor reading achievement, many

states are choosing to some form of early reading assessment policy. To date, no summary of

these current practices exists.

This study aims to explore the reading assessment practices. One purpose is to provide a

means of spreading news about successful reading assessment policies and likewise share lessons

learned to help others avoid pitfalls. A second purpose is to provide a "snapshot" of these

current practices in the reading assessment field.

In a few days you will receive in the mail a request to fill out a brief questionnaire

regarding early-grade reading assessment. You have been selected to participate because you are

either a reading or an assessment specialist within your state department of education. If you

believe that you have received this mailing in error, please contact me by electronic mail:

chrisg@wsu.edu.

I am writing you in advance because I know that people appreciate knowing they will be

contacted ahead of time. Please note that it should take about 10 - 15 minutes to fill out the

questionnaire. To facilitate your response, perhaps you could familiarize yourself with the early-

grade reading assessment practices within your state.

I encourage you take the opportunity to contribute to the reading assessment dialogue by

participating in this nationwide study. Your response is vital to improving the reading

assessment initiative.

Sincerely,   Christina L. Gilchrist

POSTCARD FOLLOW-UP

A few weeks ago a survey about early-grade reading testing was sent to you at this address.

If you have already completed and returned the questionnaire, please accept my sincere thanks. If not, please do so today. I hope you will share your experience and expertise with other states across the nation.

If you did not receive the survey, or if it was misplaced, please download a copy from http://www.wsu.edu/~chrisg/survey.pdf . Or call me at 509-332-8371 and I will mail you another today.

Sincerely,

Christina L. Gilchrist

Washington State University

Dept. of Ed. Lead. & Counseling Psych.

PO BOX 642136

Pullman, WA 99164-2136

FOLLOW-UP E-MAIL

Dear:

Last week I invited you to participate in a national study about early-grade reading assessment practices. The questionnaire enclosed has been designed to collect information about your state's policies and practices regarding early-grade reading assessment. The information you provide may help other states identify current trends and construct sound policies.

Results will <u>only</u> be summarized by state. Although I must report that I have selected participants who are either reading or assessment specialists from each State Department of Education, I assure you that your answers will be kept **strictly** confidential. Once I receive the coded questionnaires, the list of names will be destroyed. Your name will **not** be associated with the findings in any way. Remember that this survey is voluntary; if you prefer not to respond to any question you may leave it blank.

Please fill out the questionnaire at your earliest convenience. Then return it in the stamped envelope provided. A timely response from each state can enable better dialogue among the states.

If you have any questions about this study, please call me at (509) 332-8371. I would be glad to discuss any concerns you may have. This research has been reviewed and found acceptable by the Department of Educational Leadership and Counseling Psychology, the Institutional Review Board, and the Office of Grant and Research Development of Washington State University, which can be reached at (509) 335-9661.

Thank you for your time and effort in this endeavor.

Sincerely,   Christina L. Gilchrist

E-MAIL REPLACEMENT SURVEY

I was given your name as potential source for reading testing information. The questionnaire attached has been designed to collect information about your state's policies and practices regarding early-grade reading assessment. The information that you provide may help other states identify current trends and construct sound policies.

Results will <u>only</u> be summarized by state. Although I must report that I have selected participants who are either reading or assessment specialists from each State Department of Education, I assure your answers will be kept **strictly** confidential. Once I receive the coded questionnaires, the list of names will be destroyed. Your name will **not** be associated with the findings in any way. Remember that this survey is voluntary, if you prefer not to respond to any question you may leave it blank.

Please fill out the questionnaire and either mail it to Washington State University, Department of Educational Leadership and Counseling Psychology, Attn: Christina Gilchrist, PO Box 642136, Pullman, WA 99164-2136 or attach it to an e-mail at chrisg@wsu.edu at your earliest convenience. A timely response from each state can enable better dialogue among the states.

If you have any questions or comments about this study, please call me at (509) 332-8371. I would be glad to discuss any concerns you may have. This research has been reviewed and found acceptable by the Department of Educational Leadership and Counseling Psychology, the Institutional Review Board, and the Office of Grant and Research Development of Washington State University, which can be reached at (509) 335-9661.

Thank you for your time and effort in this endeavor.

Sincerely,   Christina L. Gilchrist

PHONE SCRIPT

Hello, my name is Christina Gilchrist, and I am a graduate student with the Washington State University Department of Educational Leadership and Counseling Psychology. I am collecting data for my (her) research project on Statewide Early Grade Reading Practices, and I'd like to ask you for your help by answering a few questions for me regarding *your state's testing practices.* Your participation in this survey should take about 5-10 minutes.

These data will be strictly confidential and I will not record your name. Also, your participation is completely voluntary. You are free to not answer any questions you may find objectionable, and may withdraw from the study at any time without penalty, just by letting me know you would not like to continue any further. This research has been reviewed and approved by the Institutional Review Board at WSU. If you have any questions or concerns about this research project, you can contact the WSU IRB at (509) 335-9661 or Christina Gilchrist at (509) 332-8371.

Are there any questions about my study that I can answer for you at this time? (answer questions). Would you like to participate in my study?

APPENDIX B

SURVEY

State Early-Grade Reading Assessment Practices

Directions: Please answer each question by either marking with an ×, circling your responses, or filling in the blank. When prompted by a ➔ please answer the subsequent clarifying question to the right; otherwise continue with the questions in numerical order. Feel free to use the blank sheet provided for comments.

Assessment Choice

1.   **Does your state require *statewide* assessment of early-grade (or kindergarten through 3rd grade) reading?**

   ☐   Yes
   ☐   No➔

          If there are **no current** plans for early-grade reading testing, do plans exist for early-grade reading testing in the **future**?

               ☐   Yes➔Please continue with question #2.
               ☐   No➔ Thank you for your assistance. Please return this survey in the self-addressed envelope provided.

2.   **Which of the following issues was most important in deciding to implement statewide assessments of early-grade reading? [Please mark 1 next to the description of most important issue, 2 next to the next important issue, and continue through 5 for the least important issue.]**

   a._____ Summarize reading achievement
   b._____ Predict reading achievement
   c._____ Refer students to services
   d._____ Screen for eligibility for placement into special programs
   e._____ Accountability

3.   **Which of the following policy mechanisms were used to implement the early-grade reading testing? [Please check all that apply.]**

   a.☐   Simple Information Dissemination
   b.☐   Financial Incentives/Funding Earmarks
   c.☐   Administrative Regulations
   d.☐   Legislative Regulations
   e.☐   Creation of Other Structure
   f.☐   Setting Priorities That Support Early-Grade Reading Assesssment

4.   **When did the official state requirement begin? [Please format in Month/Year]**

   _____/_____ (Month/Year of Requirement)

5. How <u>involved</u> were each of the following groups in <u>deciding</u> which assessment should be used *statewide*? **[Please circle the level of involvement for *each* group listed.]**

	Very Involved	Somewhat Involved	Minimally Involved	Not Involved	Don't Know
a. State Officials:	☐	☐	☐	☐	☐
b. School Districts:	☐	☐	☐	☐	☐
c. Teachers:	☐	☐	☐	☐	☐
d. Parents:	☐	☐	☐	☐	☐
e. Other:_____	☐	☐	☐	☐	☐

6. Why was this assessment chosen? [Please describe.]

‾‾‾‾

7. Which curricular/instructional reading philosophy does your state promote? [Check ONE.]

a.☐ Phonics
b.☐ Whole Language
c.☐ Balanced Phonics/Whole Language
d.☐ Other➔ Please Describe:_____
e.☐ No reading philosophy is promoted by my state

8. How much influence did the state reading philosophy have on choosing the assessment?

a.☐ A lot of Influence
b.☐ Some Influence
c.☐ Very Little Influence
d.☐ No Influence
e.☐ Don't Know
f.☐ Not Applicable

Assessment Characteristics

9. What is the full name(s) of the assessment(s) used? [No acronyms, please.]

‾‾‾‾ (Assessment Name)

10. Who is the author of the assessment? [No abbreviations, please.]

‾‾‾‾(Author)

11. What is the name of the publisher? [No abbreviations, please.]

‾‾‾‾(Publisher)

12.   What year was the assessment published?

_____(Year)

13.   How many forms of the assessment are available?

_____ (Number of forms)

14.   How would you describe the type of assessment? [Please check one.]

a.☐   Criterion-Referenced
b.☐   Norm-Referenced
c.☐   Other: ➔   Please describe:_____

15.   Is a technical report or manual available concerning how the assessment was developed?

a.☐   Yes➔   If yes, please describe:_____
b.☐   No

16.   Which of the following reliability measures are available? If an actual measure is available for your test, please write the coefficient in the line provided (i.e. r = +0.63).

a.☐   Test-Retest➔   Coefficient:_____
b.☐   Parallel-Forms➔   Coefficient:_____
c.☐   Split-Half ➔   Coefficient:_____
d.☐   Internal Consistency➔ Coefficient:_____

17.   Which of the following validity measures are available?

a.☐   Face Validity ➔   Please describe:_____
b.☐   Content Validity➔   Please describe:_____
c.☐   Predictive Validity➔   Please describe:_____
d.☐   Concurrent Validity➔ Please describe:_____
e.☐   Construct Validity➔   Please describe:_____

18.   The assessment requires which of the following assessment techniques? [Check all that apply.]

a.☐   Observation of student performance
b.☐   Student Product (not pencil-and-paper)
c.☐   Pencil-and-paper tests
d.☐   Anecdotal records
e.☐   Other   ➔   If "Other," please describe:_____

19.   Which of the following reading skills or behaviors are assessed? [Check all that apply.]

a.☐   Oral reading skills
b.☐   Comprehension skills
c.☐   Comprehension strategies
d.☐   Word-attack skills
e.☐   Word recognition skills
f.☐   Reading Rate
g☐   Other   ➔   If "Other," please describe:_____

20. **Are alternate forms of the assessment available?**

    a.☐   No
    b.☐   Yes ➔   If "Yes," please describe:_____

Administration of Assessment

21. **At what grouping level are students assessed? [Check only one.]**

    a.☐   Individual
    b.☐   Group

22. **How many times is the assessment administered within one school year?**

    _____ (Number of times administered)

23. **Which month(s) of the year is the assessment administered? [Please list each.]**

    _____ (Month Names)

24. **Under which conditions are kids assessed? [Circle ALL that apply.]**

	Mandatory	If Parent Requested	If Teacher Requested	If Counselor Requested	OTHER:
a. Kindergarten:	☐	☐	☐	☐	_____
b. First Grade:	☐	☐	☐	☐	_____
c. Second Grade:	☐	☐	☐	☐	_____
d. Third Grade:	☐	☐	☐	☐	_____

25. **Which part(s) of the assessment is/are used?**

    a.☐   Entire Assessment
    b.☐   Part(s) of the assessment➔ If only part is used, which parts?
                           _____ (Assessment Parts)

26. **Which mode(s) of administration is/are used? [Check all that apply.]**

    a.☐   Students read instructions individually and take test
    b.☐   Test administrator reads standardized instructions to group
    c.☐   Test administrator reads standardized instructions to individuals
    d.☐   Instructions for taking test and test itself are presented on computer screen
    e.☐   Other➔ If "Other," please describe: _____

27. **How many minutes does the assessment last?**

_____ (Number of Minutes)

28. **Who administers the assessment? [Check all that apply.]**

a.☐   Teachers
b.☐   Paraprofessionals
c.☐   District Personnel
d.☐   Other   ➔If "Other," please describe: _____

Scoring and Interpreting Assessment Information

29. **Who, or what agency, scores the assessment?**

_____ (Agency Name)

30. **What types of scores area available? [Check all that apply.]**

a.☐   Raw Score
b.☐   Percentile Rank
c.☐   Stanines
d.☐   Standard
e.☐   Grade Equivalent
f.☐   Percentage of Students Meeting Minimum Requirement
g.☐   Other➔ If "Other," please describe: _____

31. **How are scores summarized? [Check all that apply.]**

a.☐   Individual
b.☐   Classroom
c.☐   District
d.☐   State
e.☐   Other➔ If "Other," please describe: _____

32. **Who develops the report? [Check all that apply.]**

a.☐   Outside Agency➔   If "Outside Agency," please give name: _____
b.☐   State representatives
c.☐   School district representatives
d.☐   School administrators
e.☐   Psychologist
f.☐   Counselor
g.☐   Teachers
h.☐   Other➔   If "Other," please describe: _____

Assessment Results

33. **How are results communicated? [Check all that apply.]**

    a.☐   Parent-teacher conference
    b.☐   Press release
    c.☐   Special meeting
    d.☐   Other➔   If "Other," please describe: _____

34. **Who is expected to use the assessment results? [Check all that apply.]**

    a.☐   State Officials
    b.☐   School Districts
    c.☐   School Administrators
    d.☐   Counselors
    e.☐   Psychologists
    f.☐   Teachers
    g.☐   Parents
    h.☐   Students
    i.☐   Other➔   If "Other," please describe: _____

35. **How much <u>training</u> did each of the following groups receive regarding the appropriate interpretation and <u>use</u> of the assessment results? [Circle the appropriate amount of training for each group listed.]**

	Thorough Training	Some Training	Minimal Training	No Training
a. State Officials:	☐	☐	☐	☐
b. School Districts:	☐	☐	☐	☐
c. School Administrators:	☐	☐	☐	☐
d. Counselors:	☐	☐	☐	☐
e. Psychologists:	☐	☐	☐	☐
f. Teachers:	☐	☐	☐	☐
g. Parents:	☐	☐	☐	☐
h. Students	☐	☐	☐	☐
i. Others: (Please List)_____	☐	☐	☐	☐

36. How <u>informed</u> are each of the following groups about the assessment results? [Circle the appropriate level of knowledge for each group.]

a. State Officials:	Very Informed ☐	Somewhat Informed ☐	Barely Informed ☐	Not Informed ☐
b. School Districts:	Very Informed ☐	Somewhat Informed ☐	Barely Informed ☐	Not Informed ☐
c. School Administrators:	Very Informed ☐	Somewhat Informed ☐	Barely Informed ☐	Not Informed ☐
d. Counselors:	Very Informed ☐	Somewhat Informed ☐	Barely Informed ☐	Not Informed ☐
e. Psychologists:	Very Informed ☐	Somewhat Informed ☐	Barely Informed ☐	Not Informed ☐
f. Teachers:	Very Informed ☐	Somewhat Informed ☐	Barely Informed ☐	Not Informed ☐
g. Parents:	Very Informed ☐	Somewhat Informed ☐	Barely Informed ☐	Not Informed ☐
h. Students:	Very Informed ☐	Somewhat Informed ☐	Barely Informed ☐	Not Informed ☐
i. Others: (Please List)_____	Very Informed ☐	Somewhat Informed ☐	Barely Informed ☐	Not Informed ☐

Consequences

37. Are there consequences for STUDENTS based on their own performance on the reading assessment?

    a.☐   Yes➔ If "yes," please explain: _____
    b.☐   No

38. Are there consequences for TEACHERS based on students' performance on the reading assessment?

    a.☐   Yes➔ If "yes," please explain: _____
    b.☐   No

39. Are there consequences for SCHOOL DISTRICTS based on students' performances on the reading assessment?

    a.☐   Yes➔ If "yes," please explain: _____
    b.☐   No

***Thank you for your time. Don't forget to save your changes!

Please attach to e-mail to c_gil@earthlink.net
Feel free to provide additional in the body of the e-mail or in other attachments.

APPENDIX C

IRB APPROVAL

Office of Grant and Research Developm

MEMORANDUM

TO: Christina L. Gilchrist
             Educational Leadership & Counseling Psychology, WSU Pullman (2136)

FROM: Misty Cato **(for)** Michael Hendryx, Chair, WSU Institutional Review Board (3140)

DATE: 29 April 2003

SUBJECT: Approved Human Subjects Protocol - New Protocol

Your Human Subjects Review Summary Form and additional information provided for the proposal titled *"Dissertation Survey: State Early Grade Reading Assessment Practices,"*   IRB File Number   **5514**-a was reviewed for the protection of the subjects participating in the study. Based on the information received from you, the WSU-IRB   **approved** your human subjects protocol on **28 April 2003**.

IRB approval indicates that the study protocol as presented in the Human Subjects Form by the investigator, is designed to adequately protect the subjects participating in the study. This approval does not relieve the investigator from the responsibility of providing continuing attention to ethical considerations involved in the utilization of human subjects participating in the study.

This approval expires on 26 April 2004. If any significant changes are made to the study protocol you must notify the IRB before implementation.   Request for modification forms are available online at http://www.ogrd.wsu.edu/Forms.asp.

In accordance with federal regulations, this approval letter and a copy of the approved protocol must be kept with any copies of signed consent forms by the principal investigator for THREE years after completion of the project.

This institution has a Human Subjects Assurance Number M1344 which is on file with the Office for Human Research Protections. WSU's Assurance of Compliance with the Department of Health and Human Services Regulations Regarding the Use of Human Subjects can by reviewed on OGRD's homepage (http://www.ogrd.wsu.edu/) under "Electronic Forms," OGRD Memorandum #6.

If you have questions, please contact Misty Cato at OGRD (509) 335-9661. Any revised materials can be mailed to OGRD (Campus Zip 3140), faxed to (509) 335-1676, or in some cases by electronic mail, to *ogrd@mail.wsu.edu*. If materials are sent by email attachment, please make sure they are in a standard file type, (i.e., ASCII text [.txt], or Rich Text Format [.rtf]).

Review Type: NEW   OGRD No.: NF
Review Category: XMT   Agency: NA
Date Received: 10 April 2003

WASHINGTON STATE UNIVERSITY

MEMORANDUM

TO: Christina L. Gilchrist
            Educational Leadership & Counseling Psychology, WSU Pullman (2136)

FROM: Jamie Murphy **(for)** Cindy Corbett, Chair, WSU Institutional Review Board

DATE: 26 January 2004

SUBJECT: Review of Protocol Modification

Your proposal to modify the protocol titled   ***"Dissertation Survey: State Early Grade Reading Assessment Practices,"***   IRB File Number   **5514**-b was reviewed for the protection of the subjects participating in the study. Based on the information received from you, the IRB has **approved** your modification request on **26 January 2004**. This modification includes a change to conduct the approved survey over the telephone.

IRB approval indicates that the modifications described to the previously approved study protocol are designed to adequately protect the subjects participating in the study. This approval does not relieve the investigator from the responsibility of providing continuing attention to ethical considerations involved in the utilization of subjects participating in the study.

The approval for this protocol expires 26 April 2004. If any more changes are made to the study protocol you must notify the IRB and receive approval before implementation.

If you have questions, please contact the Institutional Review Board at OGRD (509) 335-9661. Any revised materials can be mailed to OGRD (Campus Zip 3140), faxed to (509) 335-1676, or in some cases by electronic mail, to *ogrd@mail.wsu.edu.*

Review Type: MOD   OGRD No.: NF
Review Category: XMT   Agency: NA
Date Received: 26 January 2004

MEMORANDUM

TO: **Christina L. Gilchrist**
Educational Leadership & Counseling Psychology, WSU Pullman (2136)

FROM: Jamie Murphy **(for)** Cindy Corbett, Chair, WSU Institutional Review Board (3140)

DATE: 24 March 2004

SUBJECT: Approved Continuing Review of Human Subjects - Continuing Review

The information provided for the continuing review of your protocol titled *"Dissertation Survey: State Early Grade Reading Assessment Practices,"* IRB Number **5514**-c was reviewed for the protection of the subjects participating in the study. Based on the information received from you, the IRB has given **approval** to continue your human subjects protocol for another year starting **24 March 2004.**

The IRB approval indicates that the study protocol as presented in the Human Subjects Form by the investigator, is designed to adequately protect the subjects participating in the study. This approval does not relieve the investigator from the responsibility of providing continuing attention to ethical considerations involved in the utilization of human subjects participating in the study.

This approval expires on 23 March 2005. If any significant changes are anticipated to the study protocol you must notify the IRB and received approval before implementation.

In accordance with federal regulations, this approval letter and a copy of the approved protocol must be kept with any copies of signed consent forms by the researcher for THREE years **after** **completion of the research.**

This institution has a Human Subjects Assurance Number FWA00002946 which is on file with the Office for Human Research Protections. WSU's Assurance of Compliance with the Department of Health and Human Services Regulations Regarding the Use of Human Subjects can by reviewed on OGRD's homepage (http://www.ogrd.wsu.edu/) under "Electronic Forms," OGRD Memorandum #6.

If you have questions, please contact the Insitutional Review Board at OGRD (509) 335-9661. Any revised materials can be mailed to OGRD (Campus Zip 3140), faxed to (509) 335-1676, or in some cases by electronic mail, to *ogrd@mail.wsu.edu.* If materials are sent by email attachment, please make sure they are in a standard file type, (i.e., ASCII text [.txt], or Rich Text Format [.rtf])..

Review Type: REN   OGRD No.: NF
Review Category: XMT   Agency: NA
Date Received: 13 February 2004