

©Copyright 2007
Patricia Marie Zurybida

Learning Together, but Differently: Understanding the Impact and Implications
of a Whole Group Reading Intervention

Patricia Marie Zurybida

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

Doctor of Education

University of Washington

2007

Program Authorized to Offer Degree:

College of Education

UMI Number: 3265441

Copyright 2007 by
Zurybida, Patricia Marie

All rights reserved.

INFORMATION TO USERS

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleed-through, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

UMI[®]

UMI Microform 3265441

Copyright 2007 by ProQuest Information and Learning Company.

All rights reserved. This microform edition is protected against
unauthorized copying under Title 17, United States Code.

ProQuest Information and Learning Company
300 North Zeeb Road
P.O. Box 1346
Ann Arbor, MI 48106-1346

In presenting this dissertation in partial fulfillment of the requirements for the doctoral degree at the University of Washington, I agree that the Library shall make its copies freely available for inspection. I further agree that extensive copying of the dissertation is allowable only for scholarly purposes, consistent with "fair use" as prescribed in the U.S. Copyright Law. Requests for copying or reproduction of this dissertation may be referred to ProQuest Information and Learning, 200 North Zeeb Road, Ann Arbor, MI 48106-1346, 1-800-521-0600, to whom the author has granted "the right to reproduce and sell (a) copies of the manuscript in microfilm and/or (b) printed copies of the manuscript made from microfilm."

Signature Patricia M. Zepeda
Date 5/31/07

University of Washington

Abstract

**Learning Together, but Differently: Understanding the Impact and Implications
of a Whole Group Reading Intervention**

Patricia Marie Zurybida

Chair of the Supervisory Committee:
Associate Professor Margaret Plecki
Educational Leadership and Policy Studies

Leaders and teachers working in schools and school districts make numerous decisions about what instructional materials or strategies to use. Often these decisions are made without collecting or examining evidence regarding the appropriateness or effectiveness of the instructional decisions. This project focuses on the *REWARDS (Reading Excellence: Word Attack & Rate Development Strategies) Intermediate* multisyllabic word reading strategies curriculum (2005) and investigates the effects of its application when taught whole class to a heterogeneous group of fourth grade students with a range of reading levels. I will be studying changes in multisyllabic word decoding accuracy and oral reading rates in all fourth grade classrooms in the district.

The increased demand for school district systems to make instructional decisions using scientifically-based practices, on-going assessment of progress and Response to Intervention (RTI) has heightened school districts' interest in assessing the growth of individual students and the rate of improvement of groups of students. Monitoring the progress of groups of students helps school teams determine not only if the intervention is effective overall, but also if the intervention is accelerating the learning rate of under performing subgroups and closing the achievement gap. The study attempts to provide a practical means for assessing student and group performance using the resources commonly available in schools and districts.

The results of this study indicate that the *REWARDS Intermediate* reading program is effective when implemented in large group, heterogeneous fourth grade classrooms. Low performing students made substantial gains in accuracy, while high performing students dramatically increased their reading rate. Middle range students made large gains in both areas.

TABLE OF CONTENTS

	Page
List of Figures.....	ii
List of Tables.....	iii
Introduction.....	1
Informing Literature and Framing Ideas.....	5
Methods.....	22
Results and Discussion.....	31
Conclusion and Implications for Leadership.....	50
References.....	57

LIST OF FIGURES

Figure Number	Page
1. Average change in 4th grade reading performance after REWARDS Intermediate	35
2. Average change in 4th grade reading performance after REWARDS Intermediate using untaught words	40
3. Oral reading rate increases for the whole group	43
4. Oral reading rate increases for benchmark students	44
5. Oral reading rate increases for strategic students	46
6. Oral reading rate increases for intensive students	47

LIST OF TABLES

Table Number	Page
1. Instructional Subgroup Definitions.....	16
2. Main Study Questions.....	23
3. Outline of Study Procedures.....	26
4. Characteristics of Schools in Hoquiam.....	29
5. Average Change in 4 th grade Reading Performance 9 weeks post-instruction.....	37

ACKNOWLEDGEMENTS

I once believed that a dissertation was the fruit of the author's labor, intelligence, and dedication. Now I know differently. While this document is my own work and required time, toil, and tenacity, at least some of the effort and struggle was shared by others. Now, three years, two babies and one doctoral program later, I hope all those who have supported me will join in celebrating the joy of this accomplishment.

I would especially like to thank my esteemed advisor, Dr. Marge Plecki, for her time and commitment. Her guidance, encouragement, and intellectual challenges were invaluable. I would also like to thank the other members of my dissertation reading committee, Drs. Tim Standal and Brad Portin for their guidance and support leading to the completion of this project.

Over the years I have been blessed to have the support of the faculty and cohort members of Danforth XIV and the Leadership of Learning Cohort 2. I am a better educator, a better leader, and most of all a better person because of all of you. A special debt of gratitude belongs to Drs. Kathy Kimball and Margery Ginsberg, who offered unceasing support, boundless confidence, and flexibility even with the joyful, but complicating, addition of children. When I doubted, they never wavered.

I have had the support of many colleagues, friends and family members. My parents, Bob and Joey Wheeldon, instilled in me a respect for learning and love of reading. Dr. Di Barr-Cole could always be counted on for a moment of sanity either professionally or personally and her recent experience navigating these waters was priceless. The dedicated staff of the Hoquiam School District, particularly former superintendent Dr. Tim McCarthy provided inspiration and encouragement. Dr. Roberta Wyde offered early counsel that made all the difference in my completion of this dissertation. In addition, I am grateful to the numerous family members and friends who have provided me with emotional or practical support over the past several years.

Finally, this dissertation would not have been possible without Paul Zurybida, my husband and friend. His support and encouragement has seen me through tumultuous times. For all those early mornings and weekends when you took extra "baby duty" so I could attend class or write, I thank you. I am also grateful for Elizabeth Zurybida and her sister, Baby Z. The two of you will never know the adjustments you had to make to accommodate my efforts, but your being provided me with more sources of inspiration around literacy, leadership, and education than I could have ever imagined before beginning this work.

Introduction

Imagine you are in a spacious, well-lit and nicely appointed conference room. Everyone seated around the polished oak conference table is dressed in suits with pens in hand, and the screen in the front of the room displays a PowerPoint presentation. Each speaker's arguments are supported with compelling numbers and graphs, the type of information most valued by these analytical minds. Yet, underneath all the trappings, there is an electric current of commitment and passion about the issue that belies their professional demeanor.

Now, move yourself to your local prison. Inmates wearing regulation uniforms are seated in a Spartan classroom under the glare of harsh fluorescent lights. Instead of an oak conference table, they sit in hard plastic chairs with attached desk tabletops. They too, have pens in hand and are also engaged in fervent and lively debate. You can hear the pain and passion within the personal accounts. In both of these contexts, the powerful and the powerless are talking about the exact same issue: illiteracy and its multiple effects on people's futures and on the broader society.

Illiteracy is recognized as having significant impacts on health and poverty. In April of 1998, Dr. G. Reid Lyon testified before the U.S. Congress Committee on Labor and Human Resources and declared that, "the National Institute of Child Health and Human Development (NICHD) within the National Institutes of Health (NIH) considers reading failure to reflect not only an educational problem, but a significant public health problem as well" (Hall & Moats, 1999). People who are functionally illiterate often are unemployed or have non-living wage jobs. Yet, the effects of poor reading do not end just at income level. Compounding the problem is a lack of health care and increased exposure to environmental hazards associated with low-level jobs and under-employment. Even simple tasks like following the directions on a bottle of cough syrup or keeping track of a child's immunization schedule are complicated by poor reading skills. In addition, approximately 70% of prison inmates scored in the lowest two levels of prose, document, and quantitative subtests of the National Adult Literacy Survey

(NALS). This led the NALS investigators to conclude, “These prisoners are apt to experience difficulty in performing tasks that require them to integrate or synthesize information from complex or lengthy texts or to perform quantitative tasks that involve two or more sequential operations and that require the individual to set up the problem” (Haigler *et al.*, 1994).

These statistics provide examples of the long-term impact of illiteracy on the lives of adults. However, these conditions are not immutable. We can, and must, intervene early so that every child is literate and can enter adulthood with the reading skills required for a successful life. In addition to alleviating some of the dismal social and financial outcomes outlined previously, the tremendous psychological and emotional costs of reading failure will be addressed as well (Snow *et al.*, 1998). This study considers one reading intervention that is designed to help children overcome what is a major hurdle for some learners: specifically the step from decoding short words slowly to decoding multisyllabic words fluently. In addition, the study addresses leadership actions that school staff can take to implement a practical means for assessing student and group performance in reading by using the resources commonly available in schools and districts.

The specific issue being examined is the use of the *REWARDS Intermediate* multisyllabic word reading program and the impact this program has on fourth graders’ accuracy and oral reading rate. Accuracy (words decoded correctly) and rate (words correct per minute) will be measured before instruction, immediately after instruction, and approximately nine weeks after instruction to look for the sustainability of the effects. Students’ performance will be reviewed as whole group and also as subgroups divided by entry reading skills. The purpose of disaggregating the data for these subgroups is to consider whether the program is equally effective, or even sufficiently effective, for all students when compared to the amount of instructional time invested. In other words, if certain subgroups are not benefiting, perhaps their time should be spent on more profitable literacy activities.

At the same time, this project is not just about one isolated reading program; it is really about social justice. Part of the responsibility of public education is to ensure that all students leave school able to read well enough to secure a living wage job, be an informed and active participant in the democratic process that guides our nation, and be a responsible global citizen. This ambitious goal is approached through a practical demonstration of one way to assess the effectiveness of instruction for a group of students. It is about helping schools and districts measure the effects of their interventions using the human, curricular, and technological resources typically available in a school. If districts can measure the impact of what they are doing, they will be better equipped to make instructional changes that will accelerate student growth. Leaders have an important role in helping districts address social justice issues such as ensuring equity of outcomes in literacy. One particular schema for analyzing leadership presented by Bolman and Deal (2003) is described in detail in the next chapter. This model is then applied to the Hoquiam School District, the setting for this study, to help explain the some of the critical decisions and actions that occurred prior to and in support of the procedures used to monitor student growth.

Organizational Overview

This dissertation is organized to first guide the reader through a description of the problem of practice and the rationale for undertaking this work and is provided in the introduction section immediately preceding this organizational overview. The next chapter includes informing literature and framing ideas around leadership, reading acquisition, and monitoring students' progress.

The third chapter describes specific methods and analytic processes that were employed in this study. Each tool used for assessment and instruction is described. In addition, an overall picture of the school district demographics and fourth-grade study participants is included. Once again, part of the purpose of this study is to offer a practical way for other

schools or districts to conduct their own research into the effects of their reading intervention programs on the whole group and specific instruction subgroups.

Results and discussion follow in the fourth chapter. The changes in student performance are reported for accuracy and oral reading rate over the three phases of the study: the five-week instructional period; the nine-week post-instruction period; and the overall fourteen-week duration of the study. This study was particularly interested in measuring the effects of the intervention on all students; each phase's results are described for the whole group and the three instructional subgroups. Finally, changes in oral reading rate for the whole group and subgroups were also compared to previously published "ambitious" and "expected" rates weekly rate increases.

Finally, the fifth chapter offers conclusions and implications for leadership. These are described using five key leadership activities: acknowledging the current reality, marshalling resources, championing implementation, reviewing results, and modifying current practice. This chapter looks at the results of this study from a systems-level leadership perspective and, ideally, offers leaders some suggestions for framing the work in their schools and districts if they want to undertake a similar process in their own educational settings.

Informing Literature and Framing Ideas

There are two strands of inquiry in this project and both have systems level leadership implications. The first strand was concerned with the effects of instruction using a particular set of materials, *REWARDS Intermediate*, adopted by the Hoquiam School District. As explained in the introduction, there are research studies that validate the effectiveness of REWARDS with intermediate level readers as an intervention tool used in a small group with struggling readers who are reading below grade at the 2.5 to 5.0 grade level and at a rate of at least 65 wcpm. However, Hoquiam made a systems level decision to teach this intervention program to a much larger group, the entire class, and with a wide range of reading skills. This decision was made based on the anecdotal observations of teachers and administrators about the progress of struggling students after using this program and the intuition that average or high performing students would also benefit from learning a reliable strategy for decoding multisyllabic words. While decision-making based on intuition is not unheard of in schools, the Hoquiam School District wanted confirmation of the effectiveness (or lack thereof) of this practice. Having solid, system-wide data on an intervention can better equip teachers, principals and central office administrators to determine if this is a useful way for all students to spend valuable instructional time and if fourth grade is an appropriate age for this type of intervention. Furthermore, superintendents must justify the curriculum and its related expenditures to the school board, who in turn, are accountable to their constituents. For this reason, it is critical that system-level educational leaders (such as curriculum directors, special education directors, and superintendents) have compelling evidence to support their recommendations.

The second strand involved inquiry from a broader perspective. It is not unusual for a teacher, a school, or a district to make a large-scale instructional decision based on little or no information about its effectiveness. The absence of data that monitors student progress to judge the impact of this decision on student learning is an even more common circumstance in schools and school districts. To further complicate the issue, there are numerous reading programs and resources that claim to be supported by scientifically

based research. The reasons for making or continuing to implement uninformed choices usually include a lack of time, money, knowledge, training, and people needed to conduct an in-depth study within a school system context. The assessment and analysis methods used in this study may have the potential to be applied to almost any reading intervention program, across a wide variety of grade levels, and with resources (time, money, knowledge, training and people) that are usually available in a typical school system. This study offered the chance to develop and test a practical system for evaluating many kinds of reading interventions using typical school resources as a means to gather evidence to inform future decision-making.

The review of foundational literature in reading acquisition begins with an introduction to reading, including the purpose and the main components of reading. Decoding and fluency will be examined in greater depth than other components because of the pivotal role these two concepts play in this study. The next section deals specifically with the Hoquiam School District and its Levels of Intervention model. It discusses how the concept of on-going progress monitoring became a critical concept in shaping the district's vision, values, and culture. This leads into the final section, leadership, where developing the culture of progress monitoring is described in terms of the specific leadership actions taken using Bolman and Deal's (2003) four frames of leadership.

Purpose of reading

There is little argument against learning to read as an important life skill. Thomas Jefferson considered the needs of society as a whole when he stated, "Democracy has no orthodoxy, but it can survive and flourish only with a literate citizenry." At an individual level, others insist that children cannot reach their full potential as adults if they are not literate (Hall & Moats, 1999). The National Institute of Child Health and Human Development (Hall & Moats, 1999; Snow et al., 1998) categorized illiteracy as a public health issue that has a negative effect on the quality of life. Quality of life is compromised through lower wages, loss of access to critical information regarding safety, health and opportunities for self-improvement, and the limitations placed on personal enrichment through learning about the world beyond one's immediate

experience including other cultures, politics, history, science, and art. This problem continues to expand in significance in proportion to increases in the pace of the information age.

Research tells us that students read less when reading is difficult (Stanovich, 1986). Therefore, these students do not develop background knowledge at the same rate as their peers. Added to this is the issue of reading rate and the fact that poorer readers read slower than skilled readers. Therefore, in the same amount of time, the quantity of words read can vary significantly. Then, a vicious cycle begins because the reading gets more difficult and the struggling student has less vocabulary and background knowledge to apply to the reading, making it even more challenging, so those students read less and fall farther behind. This is often referred to as the Matthew Effect, that is the rich get richer and the poor get poorer (Stanovich, 1986). However, systematic, explicit, structured, sequential and intense reading instruction, with a particular emphasis on early intervention, can keep the Matthew Effect from being a foregone conclusion.

Early reading acquisition

A vast amount of research and instructional materials have informed ideas about early reading acquisition (Adams, 1990; Hintze *et al.*, 2003; Langdon, 2004; McNamara *et al.*, 2005; Perfetti, 1984a, 1984b, 1986; Perfetti *et al.*, 1987; Snow *et al.*, 1998; Stanovich, 1986). Biomedicine shows promise for helping understand the synaptic action in the brain typically involved in reading. Scientists are now able to identify abnormal patterns of brain activity that are indicative of reading difficulties. Investments at the federal level have been targeted at improving understanding of early reading acquisition and ways to assist students who are struggling (Shaywitz, 2003). Students who may be at risk of having reading difficulties can be identified at an early age using relatively simple markers that appear to have a significant relationship to later reading difficulties.

Early reading acquisition is the foundation upon which advanced reading instruction is based (Snow *et al.*, 1998). Without the knowledge of the underpinnings of learning to read, we would never have discovered the importance of swift and effective intervention.

Once educators knew how to increase reading success, they applied the same principles to higher levels of reading instruction. Mastery learning and explicit instruction in sound-symbol relationships are two of the essential transferred components which appear to be critical to the success of advanced decoding instruction (Vachon & Gleason, 2004). Clearly, reading is an essential skill and the foundation of this skill lies in early reading decoding including: mastering the sound-symbol relationships; pronouncing common consonant and vowel combinations; and blending. Without the basic ability to read the words, there is no possibility of comprehending their meaning. Research studies have shown that the most pervasive and debilitating problems in reading stem directly from insufficiently or incorrectly developed decoding skills (Adams, 1990; McNamara *et al.*, 2005; Perfetti, 1986).

Effective decoding instruction teaches students to uncover the letters and patterns in words first to form an approximation of the word and then encourages informed or strategic guessing to confirm or adjust pronunciation based on context (Adams, 1990; Snow *et al.*, 1998). Experiments tracking eye movement during reading demonstrate that proficient readers look at each word and almost every letter within the word (Rayner, 1985). In other words, the brain processes letter by letter to recognize patterns first and then uses context clues as a self-check for comprehension (Cunningham, 1998; Rayner, 1985; Stanovich, 1986).

Rationale for Learning Multisyllabic Words

The topic of early decoding development has been investigated extensively (Adams, 1990; Hasbrouck & Tindal, 2004; McNamara *et al.*, 2005; Perfetti, 1984a; Perfetti *et al.*, 1987; Shaywitz, 2003; Snow *et al.*, 1998; Stanovich, 1986). However, the same level of attention has not been paid to more advanced decoding strategies. While the steps for forming sound-symbol relationships and blending the sounds into meaningful words are well researched for early reading decoding, there is less information about how students move to blending the sounds of much larger, multisyllabic words. Yet, developing these skills is critical for success in the intermediate grades of school and beyond. "Facility with big words is essential for students as they read, write and learn in all areas of school

and life. Many big words occur infrequently, but when they do occur they carry a lot of the meaning and content of what is being read” (Cunningham, 1989, p. 189). In fact, one study found that beginning in fifth grade, average students encounter approximately 10,000 novel words in their reading each year (Nagy & Anderson, 1984). In the same vein, Cunningham (1998) showed that most of these new words have at least two syllables and many have four or more. Therefore, effective decoding strategies to use on longer words are imperative for students’ content area comprehension and vocabulary development.

Unfortunately, poor decoders struggle with multisyllabic words (Feifer & De Fina, 1998; Samuels, 1988). Simply being able to decode single syllable words is insufficient to make the leap to decoding longer words (Cunningham, 1998). In fact, decoding accuracy with multisyllabic words increases the qualitative differences between proficient and weak readers (Perfetti, 1984a, 1986). The error patterns of low decoders show that these students: 1) frequently omit syllables or ignore large portions of the word; 2) mispronounce affixes; and 3) make more mistakes when reading vowel sounds than strong decoders (Cunningham, 1998; McNamara et al., 2005).

When the amount of evidence linking solid decoding skills with proficient reading is considered, it would be natural to assume that decoding skills are taught to mastery level throughout the elementary grades and beyond, if needed. Just the opposite is frequently true. Virtually all systematic, explicit decoding instruction happens in the primary grades. One of the most influential reports of its time, “*Becoming a Nation of Readers (Anderson et al., 1985)*,” stated that students did not need and should not have phonics instruction beyond the end of the second grade, except in cases of significant disability. Following this recommendation usually means that students are never instructed in strategies for multisyllabic words as the content demands increased in fourth grade and above. Instead, there is a chance that some students plateau at this level because they continue to use single syllable decoding strategies that are not effectively generalized to longer words. This study offers the opportunity to extend the science of reading development by examining a program that purports to assist students over a more advanced reading hurdle

that has not been as well researched as early reading instruction (Cunningham, 1998; Snow et al., 1998). The next section will discuss some methods that have been tried to assist students in mastering multisyllabic word decoding.

Strategies for Learning Multisyllabic Words

When students in intermediate grades and older are taught strategies for decoding longer words, their decoding and comprehension improve. A study by Lentz and Hughes (1990) demonstrated that teaching a seven step decoding strategy resulted in increased oral reading accuracy as well as comprehension for middle school students. Another study showed significant improvement in intermediate students' ability to read long words correctly after students were instructed in affixes and vowel digraphs. Surprisingly though, knowledge of the syllabication rules taught in elementary schools has not been shown to be correlated with proficient reading (Cunningham, 1998; Lenz & Hughes, 1990). It appears that these rules do not generalize into active reading. Instead students need systematic, but flexible, decoding routines that are easy to remember and work reliably for longer words instead of traditional syllabication instruction.

Since we know from brain research that decoding is heavily influenced by pattern recognition (Cunningham, 1998; Shaywitz, 2003), another strategy would be to teach the most likely patterns. First, each syllable is always going to have a vowel or vowel grapheme. Teaching students to identify and successfully pronounce these, even without breaking the word into perfect syllables, is critical to decoding success (Archer et al., 2004). Next, 80 percent of multisyllabic words have one or more affixes, so mastering these to the level of automaticity can increase the speed and accuracy of decoding (Vachon & Gleason, 2004). Using the first two techniques almost always results in a close approximation, so the third step in the strategy teaches students to use their linguistic and morphographic knowledge combined with contextual information to adjust their pronunciation. The following segment describes the ways in which the original *REWARDS* reading program, and now *REWARDS Intermediate*, intentionally includes effective instructional routines for struggling students as well as strategies for decoding multisyllabic words.

REWARDS Instruction for Struggling Learners

Effective instruction is a key element boosting the performance of struggling students and raising achievement levels for populations that traditionally achieve at lower levels (Barr & Parnett, 2003; Davis, 1997; Hasbrouck, 2004a; Leafstedt *et al.*, 2004; Shaywitz, 2003; Shepard & Smith, 1990; Snow *et al.*, 1998; Stanovich, 1986). *REWARDS* has demonstrated effectiveness with struggling learners at the intermediate and secondary level (Archer *et al.*, 2004; Vachon & Gleason, 2004). These learners develop decoding knowledge by learning generalizable, consistent strategies for breaking longer words into manageable chunks. Decoding speed improves because the strategies are routine and work consistently. Fluency is increased through the research-validated practice of repeated readings, partnering based on approximate instructional level, and student monitoring of individual progress. In addition, both *REWARDS* and *REWARDS Intermediate* uses multiple strategies that may help close the gap between grade level expectations and current performance. These strategies include repetition, pausing or “think time”, teacher modeling, frequent student responses, and using a variety of tempos and activities for presenting and practicing the same skill (Barr & Parnett, 2003; Bell, 2004; Borman, 2002; D'Angiulli *et al.*, 2004; Davis, 1997; Gerber *et al.*, 2004; Graves *et al.*, 2004; Hasbrouck, 2004a; Leafstedt *et al.*, 2004; McPeak, 2004; Shaywitz, 2003; Shepard & Smith, 1990; Snow *et al.*, 1998). In addition, struggling learners appear to benefit from explicit, intensive instruction with frequent progress monitoring that modifies the lessons (Slavin & Madden, 1989).

The *REWARDS Intermediate* program and this investigation relate to the five pillars of reading: phonemic awareness, phonics, fluency, vocabulary, and comprehension (Hasbrouck, 2004a, 2004b; Hasbrouck & Leppien, 2004; Langdon, 2004; Snow *et al.*, 1998). The most significant pillars in this case are decoding and fluency, though vocabulary development is certainly, if not centrally, addressed. Decoding is explicitly taught using systematic strategies for breaking apart longer words. To do this, students learn affixes and practice reading these word parts until they are automatically recognized. Students also learn that every syllable must have a vowel sound. Once they

identify the vowel sounds in the non-affix segments, reading those chunks becomes a simple act of decoding a one-syllable nonsense word. Students blend the affixes and nonsense parts together, adjust to make it a “real” word, and read the word smoothly, thereby increasing decoding, fluency, and word comprehension in three reliable steps.

Even though vocabulary development is not a central outcome of this program, its significant presence and contribution to the program should not be overlooked.

Sometimes, when decoding drills and repetition are emphasized, some learners develop the mistaken impression that decoding or “word calling” is the desired outcome of reading. Of course this is untrue, particularly as students engage in content area work where the emphasis needs to shift from learning to read using reading skills in order to learn. Without the successful application of those reading skills to further learning in other academic or personal interest areas reading can quickly become a pointless exercise, and one that is not going to invite the participation of struggling readers. The *REWARDS Intermediate* program explicitly teaches students the meanings of affixes, root words, and some of the newly decoded words in an engaging and thought-provoking way that challenges students to make connections not only between different forms of the word (e.g. courage and courageous) but also between words that share the same root (e.g. encourage and discourage).

The local leadership audience for this work is the Hoquiam School District. Originally, the REWARDS intervention was designed for and validated with struggling readers when taught in a small group format. While *REWARDS* has been field-tested in several studies, these have all involved intermediate or middle school students reading below grade level (Archer et al., 2004; Vachon & Gleason, 2004). The teacher’s guide suggests fourth, fifth or sixth grade classrooms as possible target audiences. However, a review of the literature reveals no published empirical studies evaluating the efficacy of a whole class implementation of the REWARDS or REWARDS Intermediate curriculum for all fourth grade students.

Regardless, the teachers and principals in Hoquiam School District were so impressed with the gains that struggling learners made while using the program, they decided to proactively provide this instructional program to all fourth graders, ensuring that all students would have the chance to learn the strategies. Furthermore, the district made this decision without investigating the potential effects on high-performing and very low-performing students, even though this program specifies that students need to be reading at a minimum of a 2.5 grade level and 60 words correct per minute (wcpm). Hoquiam, however, is not alone in this type of curricular decision. Other classrooms, schools, and districts have made similar decisions based on limited evidence. This study intends to provide some insight into the impact of this instruction when *REWARDS Intermediate* is implemented in a whole class fourth grade natural setting. The following section addressing outlines the Hoquiam School District's journey toward instructional decision-making based on on-going assessment data.

Hoquiam School District and Levels of Intervention

While gathering specific results from a whole class fourth-grade implementation of *REWARDS Intermediate* is one of the purposes of this study, school or district leaders may also be interested in the relative simplicity of the methods used to measure student progress and determine the effectiveness of the intervention. The concept of intervention, particularly in the area of reading, is a topic of significant professional interest today. Furthermore, the focus on scientifically-based interventions and on-going progress monitoring using Curriculum-Based Measures (CBM) and Response to Intervention (RTI) has generated school districts' interest in assessing the growth of individual students and the rate of improvement of groups of students. Monitoring the progress of groups of students helps school teams determine not only if the intervention is effective overall, but also if the intervention is accelerating the learning rate of under performing subgroups. Educators need to pay particular attention to whether or not interventions are effective in reducing the achievement gap that exists between students living in poverty and their more affluent peers, as well as the gap between students from racial, ethnic, and language minority groups and their majority peers. The study attempts to provide a practical means for assessing student and group performance using the resources

commonly available in schools and districts. The next several paragraphs describe the process the Hoquiam School District went through on its way to developing a comprehensive system of assessment, intervention and progress monitoring that they call Levels of Intervention. Eventually, this led to the selection of *REWARDS Intermediate* and, ultimately, the need to determine if this program was effective for all fourth grade students.

The Levels of Intervention concept represents a coordinated effort between two district departments: Special Services and Curriculum, Assessment and Instruction (CIA). The framework was developed on a foundation of research-based and research-validated practices as described in the Washington State K-12 Reading Model (Hasbrouck, 2004b). It should also be noted that Hoquiam School District's Levels of Intervention initiative was developed with active participation from over 20% of the district's teachers. Every teacher had a copy of the draft documents for review and had a chance to give input to their Curriculum Council representative. The Curriculum Council and Leadership Team, with representation at every grade level and from every school, studied the initiative closely for over a year before finalizing the recommendation to the school board, which formally adopted Levels of Intervention in June 2005. Both the *REWARDS Intermediate* instructional program and the *DIBELS* assessment were on the district's list of recommended assessments and materials. However, both *REWARDS Intermediate* and *DIBELS*, as well as the other recommendations, came with a price. Resources were critical and managing them thoughtfully through strategic leadership actions as is described in the leadership section of this chapter, enabled the district to gather enough time, money and expertise to realize our common vision. By combining resources from a variety of federal, state, and district sources and eliminating duplication of services, the district believed that it could make better use of its limited resources to intervene faster, with greater precision, and with increased intensity to help students catch up and keep up with grade level expectations.

The Levels of Intervention concept was a prescriptive approach to providing targeted intervention with the two-fold goal of increasing student achievement and making the

best use of limited time and resources possible. The interventions were based on the results of diagnostic assessments conducted by school-based teams with support from district level administrators to determine which of the core areas (phonemic awareness, decoding, fluency, vocabulary and comprehension) were weak. The diagnostic assessments also offered insight into the specific nature of the problem, including possible instructional starting points (Hasbrouck, 2004a; Hasbrouck & Leppien, 2004; Langdon, 2004; Slavin & Madden, 1989). Using the diagnostic assessment information on an individual student, the Levels of Intervention framework guided the school team's selection of a comprehensive instructional plan to meet his or her needs.

In addition to recommending instructional materials and strategies for each level of need within the five components of reading, Hoquiam's Levels of Intervention also included an assessment matrix. The assessment matrix specified the overall screening tools and diagnostic assessments that correspond to the different areas of reading. The reading screening assessment for elementary students was the *Dynamic Indicators of Basic Early Literacy Skills (DIBELS)* (Hasbrouck & Tindal, 2004). The *DIBELS* uses one-minute timings on three pieces of text, with the median score used for analysis. Students were grouped into three categories based on their statistical likelihood of meeting the reading rate benchmark standard for that grade level by spring. Students who had an 80% chance or better of meeting the benchmark standard were called *low risk* and they are referred to hereafter as *Benchmark* students (see Table 1). Students who had a 20% chance or worse of meeting the benchmark standard were called *at-risk*. These students needed intensive levels of intervention quickly in order to accelerate their progress and help them reach the spring grade level standards. These students are referred to as *Intensive* students throughout the rest of this study. Students who fell between those numbers (80%-20%) for meeting the standard were referred to as *some risk* and needed strategic instruction designed to bring them to standard. Therefore, they are designated *Strategic* students in this document. Students who showed some risk or who are at-risk for not meeting the standard were given additional diagnostic testing to determine the specific areas of need, which then shaped the type and intensity of the interventions they received.

Table 1: Instructional Subgroup Definitions

Subgroup Name	Likelihood of Meeting DIBELS Spring Benchmark
Benchmark	Over 80%
Strategic	Between 80%-20%
Intensive	Less than 20%

The *REWARDS Intermediate* program was selected as one of three tools recommended by Hoquiam's Levels of Intervention framework to support struggling intermediate readers in the areas of decoding and fluency. Hoquiam's Levels of Intervention framework was a massive undertaking to implement a 3-tiered reading model systematically in grades K-12. The intent was to develop a common district vision of comprehensive school teams that meet the instructional needs of all learners, particularly those who struggle. The Hoquiam Levels of Intervention model implements three tiers of reading intervention in order to ensure that every student receives daily reading instruction at his or her skill level. This model addresses students from kindergarten to tenth grade in the district and can be extended through high school for students who are struggling to meet state standards. In order to design and implement the Hoquiam Levels of Intervention, as well as to prepare the foundation for this study, specific leadership actions were required. Some of these actions are reviewed in the next section.

Leadership

This discussion of leadership is not designed to represent the broad spectrum of thought and research around leadership. Rather, it selects one schema for analyzing leadership and uses it to illuminate the leadership decisions made within this particular study context that prepared the Hoquiam School District with the skills to construct a practical method for gathering and analyzing the data. Hoquiam's data analysis required leaders to ask challenging questions about on-going progress for students of all instructional levels, and, equally important, to act on the data collected.

Leaders have an important role in helping districts address social justice issues such as ensuring equity of outcomes in literacy. Bolman and Deal (2003) describe four frames of leadership: structural, human resource, political, and symbolic. This study illustrates particular leadership actions taken across all four frames that provided the foundation for this study. Without these strategic leadership decisions, school district staff would not have been prepared to collect, interpret and act upon this reading data in order to evaluate this intervention's instructional effectiveness for the whole group and subgroups. This next section describes Bolman & Deal's (2003) four frames of leadership and provides actual examples taken directly from a typical school and district setting.

When operating from the symbolic frame, leaders sometimes take the role of torchbearers who spread and sustain the focused message. In the case of the Hoquiam School District, the superintendent at the time championed the message, "We assess in order to know how our students are performing and we make instructional changes based on the information to increase their progress." He addressed this topic with passion and frequency across a wide spectrum of forums like school board meetings, the Hoquiam Development Association (business association), Rotary Club, and the beginning of the year meeting to welcome back teachers. In addition, the superintendent helped others develop leadership skills and take ownership for spreading the message when he commissioned the Director of Curriculum and the Director for Special Services to write a monthly one-page instructional letter to all staff that explained a particular aspect of the Levels of Intervention. Not only did this help continue to spread and emphasize the central message, but bringing others into leadership roles in the symbolic realm activated the human resources frame as well.

The human resource frame views leadership activities through a lens that considers the skills, interests, experience, talents and weaknesses of the people in the system when shaping the approach and specific actions undertaken. In the last paragraph, an example of distributed leadership was the superintendent helping others to develop leadership skills. Other human resource actions taken in the Hoquiam School District to support the

implementation of Levels of Intervention included using two years' worth of the district-directed professional development days around concepts and materials relating to Levels of Intervention and reading. This substantial resource allocation symbolically marked Levels of Intervention as a key focus for the district and also responded to the needs of the people in the system to be knowledgeable and skilled in the implementation of the specific assessments and materials, as well as the overall structure. Also in the training arena, the Title I teachers, in conjunction with the Director of Special Services, also made a commitment to ensure that everyone who is administering the standardized assessments has been trained and passed an initial reliability check. Finally, this is the frame most often used to determine which staff members should provide particular instructional interventions. Perhaps the Title I teacher has the experience, training, and temperament best suited to deliver the intensive, direct-instruction reading program to a group of five students, three of whom qualify for special education services in reading. Effective use of the human resources frame suggests that the Title I teacher provide the instruction while consulting closely with the special education teacher. Leaders can support this process by helping to identify staff strengths, developing a schedule that allows for coordination between teachers, and taking a creative, but legally compliant, approach to funding the Title I teacher.

However, in order for the human resources frame to work most effectively when matching people to tasks and providing professional development, Hoquiam also needed to use the structural frame of leadership to build the blueprint that guided the work. Structurally, leaders develop systems or institutional routines needed to support the vision. For example, Hoquiam developed a district-wide schedule for routine screening for all students to identify those who may be falling behind in reading. This is also likely to include not only the timeline for the assessments, but also the specific assessment tools and standardized protocols for administration. They also created a matrix of interventions that matched student grade span (e.g. K-1, 2-3, 4-5, 7-12) with diagnostically-determined need (e.g. phonemic awareness, phonics, fluency, vocabulary and comprehension) and intensity of need using a three-tier model to suggest a set of materials or strategies for remediation. For example, a double dose of *Road to the Code* is typically recommended

as a first step to help a kindergarten student experiencing mild to moderate reading difficulties in the area of phonemic awareness. On the other hand, a seventh grade student who was experiencing significant difficulties in decoding (and probably fluency, vocabulary and comprehension because of the decoding issue) would most likely be given a placement test for *Corrective Reading*. Both students would receive frequent progress monitoring to ensure that the recommended program was actually making a difference in that student's learning. However, improvement on the progress monitoring assessment alone is not enough. Another structural change that Hoquiam is in the process of implementing is defining an acceptable rate of progress to ensure that students are closing the gap between their performance and grade level standards in a reasonable amount of time given their intensity of need.

District wide acceptance (or tolerance) for Levels of Intervention did not come without a price; politics were used to address this initiative. Politics plays an important role in shaping perspectives on leadership. The resources needed to provide the training and the instructional materials initially seem to be located in the structural frame. A first glance it may appear to simply be a matter of allocating fiscal resources via a procedural structure. Yet, any systems level leader will emphatically state that determining the budget is largely a matter of negotiation concerning competing priorities given limited resources, thereby placing it squarely in the political frame. Leaders need to be willing to advocate passionately, persuasively and tirelessly within the political arena in order to establish funding for priorities. Astute leaders also recognize that sometimes it is worth spending the political capital to simply insist that certain efforts be funded to particular levels. These leaders know they will need to be flexible in other areas in order to maintain working relationships with other stakeholders or they will lose their credibility and, perhaps, their jobs. However, in their analysis of the situation, these leaders have determined that the political cost of an edict in some cases is well worth the expenditure.

Hoquiam worked within the political frame to make Levels of Intervention a reality by having a very inclusive process of decision-making about the assessments and instructional materials selected. It was also stated repeatedly from the beginning that the

first year's matrix was a living changing document that would be reviewed by a group of teachers in the spring to make improvements and adjustments. Hoquiam openly acknowledged that it was unlikely that everything on the matrix would work perfectly and reassured people that changes could be made. At the same time, staff was reminded of the common vision compelling us to take action instead of waiting for the "perfect" matrix because the commitment was to adjust instruction to ensure that students were learning. Another politically strategic move was the switch from an Administrative Team summer retreat to the development of a Leadership Team that included equal representation of teachers from all buildings. Any teacher who was available during the dates of the state's Summer Institute and the Leadership Retreat could serve, regardless of their involvement with the union, other activities (e.g. coaching), or years of teaching experience. Furthermore, teachers are encouraged to serve for two years and then rotate off to help build broader capacity for leadership in the district. Teachers and principals alike agree that this action helped heal some the scars left from an eleven day strike in 2001, rebuild trust, and establish joint ownership for the teaching and learning initiatives in the district. Finally, Hoquiam bought all the materials and ensured that they were available in the schools *before* the start of the school year in sufficient quantity to begin interventions as soon as the fall diagnostic testing was completed. Hoquiam had a history of not providing core adoption materials for the Title I or Special Education teachers, so the idea that interventions would be fully funded and well supported was highly motivating for all teachers.

Sometimes rebuilding the political capital leads directly back into the symbolic frame. That was true in Hoquiam as well. Once the initiative that required difficult fiscal or human resource negotiations began to show some results, the savvy former superintendent employed the symbolic frame to tell the story that captured some of the strengths of the group. As he prepared to leave, he reminded us, "We had a vision and it was an ambitious, challenging vision, but this team knew that it would have big measurable results for kids in reading. So, we found the money; we got the training; we rescheduled the day; and we have seen tremendous results for our students. This group always focuses on making a difference for kids." Storytelling like this helped to reaffirm

the community identity, rebuild relationships by reframing difficult situations, and renewing the commitment of everyone in the organization (Bolman and Deal, 2003).

This illustration of four frames of leadership should not be considered as definitive rules for a district to pursue their own agenda of progress monitoring to ensure effective student learning. It merely offers a further explanation of one way to analyze a setting that worked well for this district and may offer insight for others. Just as conditions in other districts will vary, so too will leadership responses. Likewise, Hoquiam has not solved all of its issues. Some students are still struggling. Occasionally, the interventions are being implemented with poor fidelity. Teachers and administrators alike feel the enormous pressure to help all students learn to read in the midst of threats of decreasing staff and other resources due to declining enrollment. There are many more challenges still left for Hoquiam. The key is, however, that a beginning has been made and that more students are receiving the help they need. To that end, in the next chapter, the specific methods for examining the effectiveness of a particular intervention are described.

Methods

The analytical research questions fell into two categories. The first category examined changes in students' reading rate and accuracy as a whole group and tried to answer questions like, "What impact does completing the *REWARDS Intermediate* program at the beginning of 4th grade have on students' ability to read multisyllabic words and on the rate at which they read grade level text?" In order to assess these changes, accuracy and fluency were assessed at the end of instruction and again approximately nine weeks later to see if gains persisted. Related research questions included: Does using the *REWARDS Intermediate* program increase the multisyllabic word decoding accuracy and oral reading rate for fourth graders when taught in a whole group context? Do any improvements in decoding and rate persist after the program is completed?

The second category of research questions dealt with the impacts of the *REWARDS Intermediate* program on instructional subgroups of students. Some of the associated questions were, "If students were assigned to reading achievement level groups, which group or groups appear to benefit the most? Do the accuracy or rate scores for any subgroup of students decrease or show only minimal gains?" Using the accuracy and fluency scores collected at the end of instruction and again approximately nine weeks post-instruction, the data were analyzed by subgroup. Table 2 below depicts the research questions and the related study strategies along with the data sources used. The *REWARDS Intermediate* pre/posttest and generalization test were lists of twenty multisyllabic words designed to measure students' accuracy in decoding word parts and whole words. The *REWARDS Intermediate* fluency measure was a one-minute timing on an expository piece of text. All of these tools were included in the *REWARDS Intermediate* Teacher's Guide and will be discussed in more detail in the next section of this chapter. *DIBELS*, the assessment used to separate students into instructional subgroups, was described in the previous section as an illustration of one of the screening assessments used in Hoquiam's Levels of Intervention.

Table 2: Main Study Questions

Key Questions	Strategy/strategies employed	Data source/sources
Does accuracy improve?	Decoding accuracy scores on multisyllabic words	<i>REWARDS Intermediate</i> pre/posttest given within two weeks of starting and completing the program
Does fluency improve?	Oral reading rate measures from a one minute timing	<i>REWARDS Intermediate</i> fluency measure given within two weeks of starting and completing the program
Do gains persist?	Decoding accuracy scores on multisyllabic words <i>and</i> Oral reading rate measures from a one minute timing	<i>Nine weeks post instruction REWARDS Intermediate</i> generalization and fluency measure
Which group of students (intensive, strategic, or benchmark) benefits the most?	Separate students into groups based on current reading performance Compare the gains made by each group using posttest and generalization scores	Fall Benchmark <i>DIBELS</i> done within the first 3 weeks of school <i>REWARDS Intermediate</i> pre/posttest, generalization, and fluency measures

This study employed quasi-experimental methods. Measuring and quantifying data in the manner described above is the best approach for answering the research question for three reasons. First, measurements must be reliable over time (Creswell, 2002). Second, it is helpful to have some comparative information. Third, by standardizing the tools and data collection procedures, reliability increased and some personal subjectivity decreased.

This study's quasi-experimental methods used tools that allowed for precise measurement of student performance at three time periods: prior to instruction, immediately following instruction, and nine weeks after instruction. These methods allowed for the collection of data on the persistence and generalization of the skill development. Precise measurements show the amount of change this program made in decoding and oral reading rates.

Standardized protocol for administering the tests increased the likelihood of inter-rater

reliability (Borg & Gall, 1983; Creswell, 2002). It also provided for a common language to use when describing changes in student performance. In addition, these measurements can be compared against other curricular materials using the same measurement protocol (Creswell, 2002; Miller, 1997). Having a standard for comparison is important so growth in specific skills can be measured in similar ways. Finally, standardizing the procedures for collecting and analyzing the data reduces personal bias that may exist when conducting the study.

Procedures

This detailed description of procedures is followed by a summary chart and timeline. In order to pursue the question of decoding accuracy, a pretest/post test procedure with multisyllabic words in isolation was used. At the end of the instructional period, approximately twenty-five school days later, students were tested once again on the same multisyllabic words using the pre/post test form that accompanies the *REWARDS Intermediate* teacher's guide. This instrument measures 20 multisyllabic words and 78 word parts. The word part measurement is more sensitive than the word measure and therefore showed finer gradations of change, a useful feature for distinguishing the amounts of growth between subgroups. This construct for basic decoding measurement is typical for similar studies and widely accepted as a reasonable method for examining change (Barr & Parnett, 2003; D'Angiulli et al., 2004; Hasbrouck, 2004a; Langdon, 2004; McPeak, 2004; Vachon & Gleason, 2004).

The demands of everyday life, however, rarely call for reading lists of random multisyllabic words. In order to make this study a more valid measure of growth in functional reading, it was important to measure rate and decoding accuracy in actual pieces of text. For this reason, students were given an individual one-minute oral reading timing of a text selection. Their performance was scored for overall decoding accuracy and rate, with rate being judged as words correct per minute. Words that were mispronounced, omitted or substituted were counted as errors. Self-corrections were allowed within three seconds of the original error. Insertions were ignored and did not count as a correct word or an error. This procedure is also commonly used and accepted

and reflects the recent work of Hasbrouck and Tindal (2004) in resetting the national standard for oral reading fluency rates (Hasbrouck, 2004a; Langdon, 2004; McPeak, 2004; Shaywitz, 2003; Snow et al., 1998; Stanovich, 1986). The pretest and posttest used the fluency measure provided in the *REWARDS Intermediate* program.

Positive posttest outcomes, however, do not give enough information about the sustainability and generalizability of the skill increases. For that, students were tested with untaught multisyllabic words at nine weeks post-instruction. The generalization accuracy measure, given approximately nine weeks after instruction, also came from *REWARDS Intermediate*. Like the pre/post test tool, it contained 20 multisyllabic words and 78 words parts; however, these were novel words not taught in the *REWARDS Intermediate* program. The oral reading rate was also measured using the *REWARDS Intermediate* fluency test. Comparing the results of the generalization test to the posttest provide data to help determine if the gains in decoding accuracy and fluency seen during the program persisted. Likewise, comparing the results of the generalization test to the pretest illustrated how well students had learned the underlying decoding strategy in a new situation. Ultimately, the true measure of the program is students' ability to apply the strategy outside of the teaching environment.

Finally, this inquiry was designed to examine which group or groups of students appear to benefit the most from the *REWARDS Intermediate* intervention. The tool selected for this portion of the study was the *DIBELS* fourth grade Fall Benchmark Assessment. Entry reading level was important in determining whether this curriculum produces substantial measurable effects for all students (Borg & Gall, 1983). Otherwise, it is questionable whether a whole class application of this curriculum is either necessary or beneficial. The effect of a whole class application of the small-group instructional intervention materials was investigated so that teachers, principals, and district administrators can use the information when making decisions about their programs and curriculum.

Table 3 below describes the specific study procedures, the amount of student time required, and the people who administered the assessments. The final column explains

which procedures are routine district practices. The final two rows of study procedures are unique to this research project.

Table 3: Outline of Study Procedures

Study Procedure	Study Phase	Student Time Required	Administered by	Is this a routine district practice?
<i>DIBELS</i> Fall Benchmark	Baseline data	3 minutes (3 one-minute timings)	Teachers and other district staff	Yes
<i>REWARDS</i> Pretest word list (20 words)	Baseline data	1-2 minutes	Teachers and other district staff	Yes
<i>REWARDS</i> Pretest fluency timing	Baseline data	1 minute	Teachers and other district staff	Yes
Instruction in <i>REWARDS</i> Intermediate materials	Instruction	25 lessons (about 20-25 hours over approximately 25 days)	Classroom teacher	Yes
<i>REWARDS</i> posttest word list (20 words)	Immediate Post-intervention data	1-2 minutes	Teachers and other district staff	Yes
<i>REWARDS</i> Posttest fluency timing	Immediate Post-intervention data	1 minute	Teachers and other district staff	Yes
<i>REWARDS</i> generalization word list (20 words)	Follow-up intervention data (about 10 weeks after instruction)	1-2 minutes	Teachers and other district staff	No, required consent and assent
<i>REWARDS</i> Posttest fluency timing	Follow-up intervention data (about 10 weeks after instruction)	1 minute	Teachers and other district staff	No, required consent and assent

The school district routinely administers and analyzes the *DIBELS* assessment. The *REWARDS Intermediate* materials are district-adopted, required curriculum for fourth grade. The pretest/posttest word lists and fluency passage are included in the materials and have been given as a part of the program.

The only new procedures that were not done previously as a part of standard district practice were the administration of the generalization word list and a third administration of the fluency passage approximately nine weeks after instruction. Both of these tools are in the *REWARDS Intermediate* materials, but they had not been used routinely before in this district. Together, these assessments took about an additional three to five minutes of student time. The combination of dividing students into skill levels using the *DIBELS* data and analyzing their progress according to these divisions was unique to this study. The students and their teachers were not aware of these divisions; they were only for analytic purposes and did not impact the instruction.

Research participants

In addition to the historical information provided in the last chapter about the Hoquiam School District and its involvement in Levels of Intervention, the demographics of Hoquiam as a district, the arrangement of the fourth grade classes, and the composition of the study participants are outlined below. Hoquiam's choice to use a systematic, district wide approach for assessing and intervening has changed the culture of the district by creating a common expectation that all students should be taught at their instructional level and the demand for assessments and instructional tools that effectively diagnose and remediate problems areas. Beyond simply having the expectation that this must happen, staff members have the confidence that they have the knowledge, skills, and support needed to assist all students in learning to read. This common vision along with the underlying structure provided by Hoquiam's Levels of Intervention made this district a good setting for understanding the effects of the *REWARDS Intermediate* reading program.

Hoquiam is a small, rural school district with just under 2,000 students. The student population is distributed across six school buildings. One school serves all three hundred of the district's students in preschool, kindergarten and first. The middle school has about 370 students in grades seven and eight. Hoquiam High School is a traditional ninth through twelfth grade structure with about 700 students. Students in second through sixth grade are served in three neighborhood schools of approximately two hundred students apiece. These small schools afford a natural intimacy for students and staff stemming from the ability to know one another. District needs assessment reports show that communication and a sense of belonging are enhanced by the small school size.

The students attending Hoquiam schools are fairly homogeneous. Eighty percent of the students are White and sixteen percent are Hispanic. The remaining four percent represents all other ethnicities. Three percent of Hoquiam School District's population is English Language Learners and all of these students are native Spanish speakers. . Hoquiam is an economically depressed community and students are also largely similar in socio-economic status. The percentage of students in Hoquiam's elementary schools who receive free or reduced price lunch ranges between 59% and 70%. While many of the people in the community have lived here for generations, the traditional livelihoods provided by fishing and logging declined rapidly in the 1970s and have continued to disappear. Initially, families were suddenly thrown into poverty, but now it has become a generational issue. Table 4 below provides demographic and program data about each of the district's schools.

Table 4: Characteristics of Schools in Hoquiam

Name	Program	Size (May 2006)	Free & Reduced Lunch rate (May 2006)	Race (May 2006)	Special Education (May 2006)	English Language Learners (May 2006)
Emerson School	Preschool, Kinder- garten & 1st grade	314	63.7%	Am. Indian 5.4% Asian 2.2% Black 0.6% Hispanic 17.8% White 72.9%	21.3%	9.2%
Central Elementary	2 nd – 6 th grades	246	60.5%	Am. Indian 8.5% Asian .8% Black 1.2% Hispanic 13.0% White 76.4%	9.3%	4.4%
Lincoln Elementary	2 nd – 6 th grades	220	70.2%	Am. Indian 9.5% Asian 1.4% Black .9% Hispanic 15.5% White 72.7%	12.7%	4.9%
Washington Elementary	2 nd – 6 th grades	185	59.7%	Am. Indian 7.0% Asian 3.2% Black 1.1% Hispanic 9.2% White 79.5%	10.2%	2.7%
Hoquiam Middle School	7 th – 8 th grades	367	59.0%	Am. Indian 8.2% Asian 3.5% Black 1.1% Hispanic 9.5% White 77.7%	12.2%	2.4%
Hoquiam High School	9 th – 12 th grades	753	46.6%	Am. Indian 7.6% Asian 3.1% Black 1.1% Hispanic 6.5% White 81.4%	12.2%	1.5%
District Total	2085 headcount	2085	57.2%	Am. Indian 7.6% Asian 2.6% Black 1.0% Hispanic 10.7% White 77.8%	13.1%	3.7%

Hoquiam was a good setting for this research study because they were using *REWARDS Intermediate* in heterogeneous fourth grade classrooms as a whole district implementation. Consequently, this provided more data than a single classroom study. In an individual classroom, a particular teacher may be more skilled at reaching some students than others and may be routinely more effective with those students regardless of what program is used. More students and classrooms reduced the impact of one specific teacher's skills and preferences on student performance.

Having multiple classrooms within the same district was also a tremendous asset to this study. Since all the classrooms responded to the same district adopted curriculums, policies, and procedures, variation across schools on these dimensions was significantly reduced. Even the quantity of teacher, paraeducator and specialist (Title I, LAP, Special Education) support was essentially uniform across the schools within the district when viewed from a per student perspective. In addition, many of the students were moving into fourth grade from the same curricular materials base. In the 2004-05 school year, the Hoquiam School District formally adopted the *Harcourt Trophies* reading series, making Harcourt the core curriculum that all elementary teachers used to teach reading. Obviously, the previous teachers of these fourth graders might have varied in their teaching skills and areas of emphasis, but the students who have been in the Hoquiam School District have had very similar curricular opportunities.

This study spanned all the 4th grade classrooms and involved 137 students with a range of reading abilities, attitudes, and aptitudes. All enrolled students were given the fall *DIBELS*, the pretest accuracy and rate measures, and the posttest accuracy and rate measures since these were a part of standard district procedure. The final assessments, given at about nine weeks post-instruction, were only given to students who had been continuously enrolled, had parent consent, and had student assent. Therefore, only 100 students had all the components required for analysis. Six students were not enrolled during the Fall *DIBELS* Benchmark Assessment window. Ten students entered the district after the pretest was administered and instruction had begun. Three students left the district prior to the posttest and two students left during the nine-week period between the end of instruction and the generalization assessment. In total, fifteen students (11%) were excluded from the data analysis due to transience. Eighteen students (13%) never returned a consent form. Most classrooms had high consent return rates (average 95%), though two classrooms located in different buildings had rates of return of 63% and 69%. The rate of return appeared to be unrelated to student achievement levels. One family (.7%) officially declined consent and two students (1.5%) refused assent. These three responses were randomly spread between classes and buildings.

Results and Discussion

Analytic Techniques

The first step in the analysis was comparing the pretest and posttest data for individual students to determine how much change the student made in decoding accuracy or reading rate for words in isolation and passages. The average change in student performance described the type of progress made by the group as a whole. Next, the results from the posttest and the generalization probe were given the same comparison to determine whether students' skills were sustained, improved, or declined when explicit instruction was discontinued. Finally, subgroup data were analyzed in the same way three more times. This provided information about which group or groups of students appear to benefit the most in accuracy and rate and to what level the whole group or subgroups sustained improvements in accuracy and rate over time.

The previously described three categories for dividing the fourth graders into subpopulations (intensive, strategic, and benchmark) were used to analyze the relative amounts of progress made by each group. Given the anticipated disparity in reading achievement levels, it was expected that the benchmark group would still read faster and more accurately than the intensive group, even after *REWARDS Intermediate* instruction. Relative comparisons of progress were critical in this case to determine if all students are progressing at the same pace, or if certain subpopulations actually accelerate their achievement beyond the other subgroups' performance levels. Ideally, the whole class implementation of the *REWARDS Intermediate* strategies would have resulted in significantly increased reading rate and decoding accuracy for all students. It was, however, even more important to note accelerated growth in the strategic and intensive populations, indicating whether or not the disparity between the original subgroups had narrowed.

Quality of Inquiry

Validity requires that one proves that what is claimed to be measured is actually what is being measured (Glicking, 2005). For example, it would not be valid to measure number sense by asking a student to sort a deck of playing cards into piles. The student may

simply be matching the symbols or number of objects on the cards. A valid measure of number sense would be to have the student count a set of objects and write the appropriate numerical symbol or to give the student a numerical symbol and have him select the corresponding quantity of objects. Employing both measures would be an even stronger method.

For this investigation two different ways of measuring decoding accuracy were employed. First, students read words in isolation. This is useful because it removes the contextual clues about a word that often times can be found in passages (Langdon, 2004). Students can only rely on their decoding strategies in this situation. In addition, words read in isolation can be scored accurately for the correct decoding of word parts. Generally, words read in text are read far too quickly and fluidly for such a discrete analysis. Given intensive training and the right technology, this level of analysis can be done. However, this study is also committed to providing a procedure for analyzing student growth that can be used with the staff and resources routinely available in school districts. The second measure of decoding accuracy is reading words in text. This allows students to use context clues, but also adds the burden of unconscious comprehension processes as the students try to make meaning from the text while reading accurately and rapidly (Hasbrouck & Tindal, 2004).

Reliability refers to the degree that, given the same conditions and protocol, the same results will be produced (Borg & Gall, 1983; Stringer, 2004). There are several tenets of reliability including: the standardization of the protocol, the uniform administration of the stimulus, and the similarities in rating among observers. The first form, a standardized protocol, was addressed in this study through the *REWARDS Intermediate* curriculum itself. It is a direct instruction and scripted program that increases the likelihood that all students will have access to the same concepts and a similar number of practice opportunities, thereby meeting the second tenet of reliability – uniform administration of the stimulus (Stringer, 2004). The assessment is also standardized. An advantage of having the pretest, posttest and generalization probe all from the same source is the consistency of format and presentation. The same reasoning applies to the text passages

for oral reading fluency. This ensures that students are not distracted or confused by changes in the format or presentation of the items. Only one person administered the *REWARDS Intermediate* accuracy and rate measures, so the issue of inter-rater reliability was moot. In exchange, the study ran the risk of intra-rater drift, the possibility that the individual administering the test would begin to change the standards over time. The standardized testing format also helped with presentation and score consistency in order to reduce this potential threat to reliability (Borg & Gall, 1983).

Structuring the instructional and assessment materials is only a portion of the reliability challenge. Individuals can still introduce variables in the assessment and instruction if they do not understand or vary from the protocol (Borg & Gall, 1983; Glicking, 2005). To address these variables, joint training and practice sessions were made available for raters. The Hoquiam School District already employed two certified *DIBELS* trainers prior to this study. Before raters were allowed to administer the *DIBELS* assessments, they demonstrated proficiency in standardized assessment and scoring procedures as outlined in the *DIBELS* administration guide. Both of these factors served to increase inter-rater reliability (Borg & Gall, 1983).

Applicability is based on whether this study could be applied in other contexts (Stringer, 2004). For example, do the results and conclusions established for the Hoquiam School District have any meaning for the neighboring, and demographically similar, Aberdeen School District? Can they be applied further, perhaps to an urban, ethnically diverse school district on the East Coast? Even more important is the protocol for comparing the results of other intervention programs in the future (Borg & Gall, 1983; Shaywitz, 2003). If district administrators, principals and teachers have a straightforward, inexpensive and fairly quick method of reliably collecting and analyzing performance data, they will have another tool for monitoring not only individual student progress, but also the impact of a particular teacher's implementation with a group of students and the overall efficiency and effectiveness of interventions. These skills could potentially be applied beyond this one program or subject area. It is not difficult to move from collecting and reflecting critically on performance data in reading to data in other subject areas, particularly those

with specific standards. Money and time are scarce commodities everywhere and sometimes schools are ready to move deeper into adjusting instruction based on assessment data before the district has adopted changes. This study develops and tests a realistic, step-by-step procedure that can be applied by a district, school or an individual teacher.

Pretest to Posttest Changes

During the pretest, the first part of the assessment asked students to read a list of twenty multisyllabic words like “combination” and “irregularity”. Students were scored on each word part, roughly a syllable, they read correctly out of a total of 78. As a whole group, students averaged reading 60 word parts correctly. The most intensive students read an average of 47 word parts. The middle range of students, the strategic group, read an average of 60 word parts, which matched the whole group’s performance. Finally, the highest achieving, or benchmark students read an average of 68 word parts correctly.

The changes from pretest to posttest in correctly read word parts were dramatic. It was hypothesized that the intensive students would make the largest gains in accuracy because the program contained such explicit instruction and practice in vowel diagraphs and affixes. As anticipated, the lowest performers made the most growth in accuracy, gaining an average of 22.4 word parts during the five weeks of instruction. Adding this increase to their pretest score of 47 equals an impressive average posttest word part score of 69, the difference between 60% accuracy and 88% accuracy. It was also anticipated that the strategic students, those in the middle range, would substantially improve their accuracy. This group gained an average of 14.29 words bringing them to 95% accuracy. The amount of progress made in accuracy by the benchmark group was unexpected, since they were already reading an average of 87% of the word parts correctly on the pretest. They showed more progress than expected and gained an average of 8.24 word parts or almost 98% accuracy. Overall, the whole grouped increased an average of 13.28 word parts or 95% accuracy. These results are also displayed in Figure 1 below.

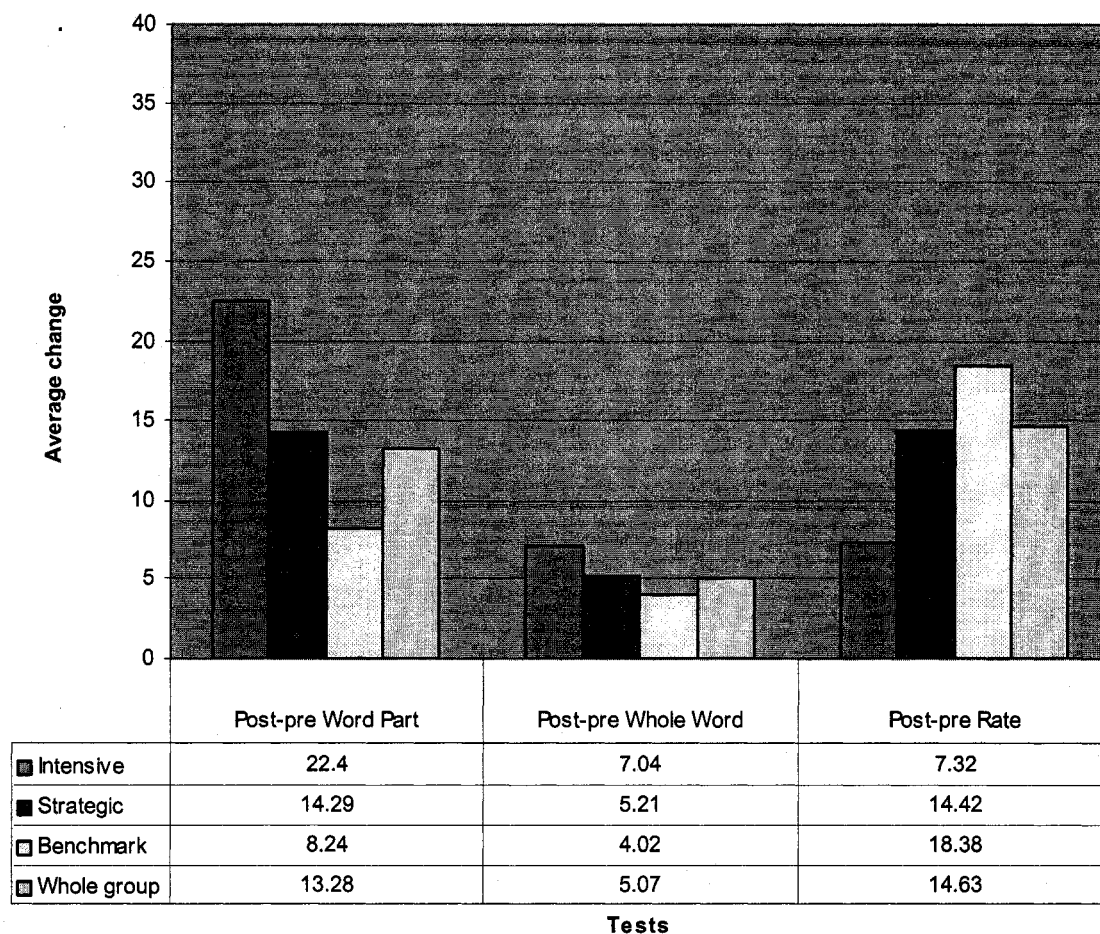


Figure 1: Average change in fourth grade reading performance after REWARDS Intermediate

In addition to scoring word parts, each student's accuracy in reading the whole word out of twenty possible words was also measured. While the changes are less striking, they mirror the changes in word part accuracy as shown in Figure 1. Intensive students gained an average of 7.04 correct whole words. Strategic students gained 5.21; while the benchmark students and whole group gained an average of 4.02 and 5.07 correct whole words, respectively.

Finally, each student was given a one-minute timing on a fluency passage. The amount of change in rate from pretest to posttest was the opposite of the pattern for accuracy. It was hypothesized that the students who were already strong decoders, those in the benchmark group, would make the largest gains. They did, gaining an average of 18.38 words correct

per minute (wcpm) during the five weeks of instruction. It was also anticipated that the strategic students would substantially improve their rate since this program was specifically targeted to help struggling readers in their skill range. This group gained an average of 14.42 wcpm. The amount of progress made in rate by the intensive group was unexpected, since they were struggling so much to simply decode the words. However, even the intensive group showed more progress than expected and gained an average of 7.32 wcpm. Overall, the whole group increased an average of 14.63 wcpm.

Posttest to Generalization Test Changes

Approximately nine weeks after instruction with *REWARDS Intermediate* was completed and the posttest was given, students were assessed again to measure how well they had sustained their ability to decode multisyllabic words accurately. Once again, students were asked to read a list of twenty multisyllabic words. While the word list was approximately that same level of difficulty as the pretest and posttest, these words had not been taught during instruction. Like the pretest and posttest, students were scored on each word part they read correctly out of a total of 78. Results are depicted in Table 5. The first two columns, Gen-Post Word Part and Gen-Post Whole Word, represent decreases in student accuracy after instruction in *REWARDS Intermediate* stopped. It is not surprising that students would be more skilled at applying the decoding routine after five-weeks of daily guided practice than after a nine-week delay. Therefore, the data in these two columns are negative, showing a decrease in decoding accuracy. The objective was to see how capable students were at retaining the decoding routine and implementing it independently on words they had never seen in the *REWARDS Intermediate* program. The third column of data represents the on-going oral reading rate increases. While it was not expected that students would suddenly begin to read more slowly, it was hypothesized that they might gain words more slowly. Therefore, the data in this set of columns is positive.

Table 5: Average Change in 4th grade Reading Performance 9 weeks Post-instruction

	Gen-post Word Part	Gen-post Whole Word	Gen-post Rate
Intensive	-2	-1.52	12.64
Strategic	-0.33	-1.00	16.79
Benchmark	-1.36	-1.08	16.76
Whole group	-1.27	-1.17	15.73

A substantial drop in students' ability to accurately decode multisyllabic words was expected because of the time elapse after instruction and because the students were not engaging in specific daily practice of the vowel digraphs, affixes, and strategy application. This decrease was expected to be quite large for the lowest performing, intensive subgroup since these students were not strong decoders at the beginning of the program. In fact, many of the students in the intensive subgroup were reading below the general entry criteria of the program, which suggests a reading grade level of 2.5 to 5.0 and a minimum rate of 60 wcpm. The assumption was that these students would need more on-going support than the five weeks of whole group instruction to maintain their ability to use the decoding strategy. In a similar fashion, no decrease or a very slight decrease in accuracy was predicted for the highest performing subgroup of students because they entered the program as skillful decoders and it was assumed that instruction with *REWARDS Intermediate* would not improve their accuracy in any significant way. Therefore, it was hypothesized that they would be less likely to lose what they had learned.

The changes from posttest to generalization test did show a drop in accuracy, but much smaller than had been projected. As anticipated, the lowest performers decreased the most in accuracy, decreasing an average of 2 word parts from 69 on the posttest to 67 on the generalization test. While this was the largest decrease experienced by any of the subgroups, it is a much smaller decrease in accuracy than originally hypothesized and

supports the assertion that even low performing students learn and retain the phonics instruction and decoding strategy taught in *REWARDS Intermediate*.

Strategic students, the ones performing in the middle range, experienced the smallest loss of accuracy, losing only .33 of a word part nine weeks post-instruction. To clarify, that is less than one word *part*, virtually no decrease at all. It appears that the decoding practice and strategy instruction were quite effective and were well sustained by this subgroup. Somewhat surprisingly, the highest performing students, those in the benchmark group, had greater declines in accuracy than the mid-range students. They lost an average of 1.36 word parts after nine weeks. Still, this only reduced their word part accuracy from almost 98% to 96%. Overall, the whole group decreased an average of 1.27 word parts.

In addition to scoring word parts, each student's accuracy in reading the whole word was also measured out of twenty total possible correct words. These changes appear to be more striking because the unit of analysis is the whole word, with a total possible of 20, rather than word parts, with a total possible of 78. They do, however, mirror the changes in word part accuracy. Intensive students decreased an average of 1.52 correct whole words. Strategic students lost 1.0, while the benchmark students and whole group declined by an average of 1.08 and 1.17 correct whole words respectively.

A one-minute fluency measure was also repeated as a part of the generalization assessment. Prior to this study, I expected students to continue to increase their oral reading rate, but at a pace much closer to their expected rate of improvement taken from the Hasbrouck and Tindal (2004) tables. Nonetheless, students continued to show strong gains in oral reading rate. Unexpectedly, the benchmark students and the strategic students had virtually identical gains in fluency increasing by an average of 16.76 and 16.79 respectively. Intensive students increased their oral reading rates by an average of 12.64 wcpm during the nine weeks after instruction. During this same time, the whole group rate increased by an average of 15.73 wcpm.

Pretest to Generalization Test Changes

After a detailed consideration of the data changes from pretest to posttest and posttest to generalization test, it might seem redundant to consider changes in student performance between the pretest and the generalization test. It is important, however, to examine the accuracy data set in order to understand how well students were able to transfer their skills in decoding multisyllabic words when presented with a novel word list. In other words, the accuracy changes from pretest to posttest gave an indication of how well the students applied their learning immediately after instruction to words that had been taught in the program. Here the students may have benefited from having seen the word before and this may have inflated their accuracy scores. The accuracy changes from posttest to generalization test were a measure of how well students could sustain their ability to apply the strategy to novel words nine weeks after instruction. In this case, students were at a double disadvantage both with time and with new words, so it is not surprising that there were decreases. However, what is really important is students' ability to generalize their learning to new multisyllabic words long after instruction ends. The comparison of accuracy from pretest to generalization test helps to answer this question. Results are displayed in Figure 2.

The changes from pretest to generalization test in correctly read word parts were encouraging. It was expected that the students who struggled the most, the intensive students, would make the largest gains in accuracy during instruction because the program contained such explicit instruction and practice in vowel diagraphs and affixes. However, it was also expected that the students in the strategic group to be the best at sustaining and generalizing their decoding strategy because they had plenty of room for growth, unlike the benchmark students. Likewise, the strategic students also already had a solid foundation in decoding before the program started making new learning easier, in contrast with the intensive students, who were still mastering basic decoding. In short, the strategic subgroup represented the students for which the program was originally designed.

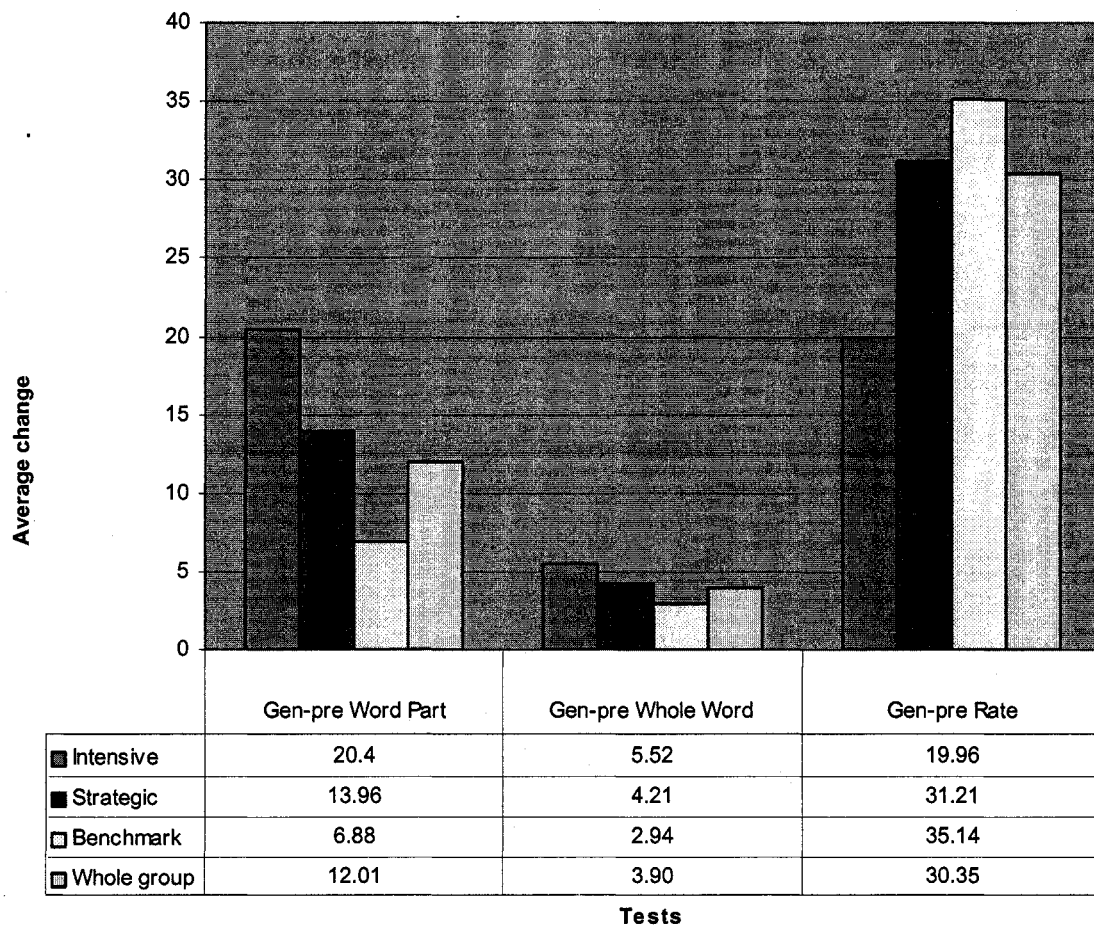


Figure 2: Average change in 4th grade reading performance after REWARDS Intermediate using untaught words

Over the entire fourteen weeks from pretest to generalization test, the intensive students made the most gains in accuracy and increased by 20.4 words correct per minute (wcpm). The strategic group gained 13.96 wcpm. The benchmark group gained 6.88 wcpm and the whole group gained 12.01 wcpm over the fourteen-week period of this study. While this matched the previous pattern from pretest to posttest, it did not match the hypothesis that low performing students would not retain their decoding skills after instruction and daily practice stopped. In fact, the results support *REWARDS Intermediate* as a more powerful way to teach decoding skills to lower performers than originally anticipated.

When whole words are considered, the intensive students gained an average of 5.52 correct whole words. Strategic students gained 4.21; while the benchmark students and whole group gained an average of 2.94 and 3.90 correct whole words respectively. This is also consistent with previous achievement patterns.

Finally, when oral reading rate is examined, intensive students gained an average of 19.96 wcpm during the fourteen weeks of instruction. The strategic students, those in the middle range, improved by an average of 31.21 wcpm. The benchmark students once again made the largest gains in words correct per minute, with average increases of 35.14 wcpm over the fourteen weeks. The average gain for the entire group was 30.35 wcpm.

Gains in oral reading rates compared to published norms

Another interesting aspect of oral reading rate changes is the comparison of actual rate changes to expected values located in the research. This study considered two informing sources. The first source developed grade level standards for “realistic” and “ambitious” goals, asserting that students who were behind needed to make better than average progress in order to catch up to their peers’ performance (Fuchs & Fuchs, 1993). The standard for fourth grade was set with .85 words gained per week as realistic and 1.1 words gained per week as ambitious. The analysis of the data from this study compares students’ progress with the ambitious goal because intervention must make headway in closing the gap between performance and grade level standards for intensive and strategic students. Furthermore, if the highest performing students make ambitious growth, then the argument that their time is being used effectively is more easily supported. Unfortunately, this research only listed one value for realistic and ambitious per grade level. It may be more helpful to consider a variety of rates of progress that more closely approximates the growth difference we see for the spectrum of students within a grade level.

The next informing source breaks out expected student performance based on the student’s current levels of achievement approximated from ranges near their percentile rank. In 2004, Hasbrouck and Tindal re-normed the oral reading rates for students in the

first to eighth grade. For each grade level, they provided statistics on average words correct per minute for students at the 10th, 25th, 50th, 75th, and 95th percentiles. In addition, they also calculated an “expected” rate of growth in words gained per week for each percentile level at each grade level given student placement in a solid core reading program. In other words, the expected rate of growth for an intensive fourth grade student in the 10th percentile (.8 wcpm/week) could be compared to the average expected rate of growth for a benchmark fourth grader in the 90th percentile (1.1 wcpm/week). In the next several paragraphs, the growth of the whole group and the subgroups is compared to their ambitious and expected rates of growth during the five-week instructional period, the nine-week post-instruction period, and the total study period of fourteen weeks.

First, the progress of the whole group was considered. In the five weeks of instruction between pretest and posttest data, the ambitious goal for increasing rate was 5.5 wcpm (see Figure 3). The average gain for the entire group was 14.63 wcpm, or 2.66 times the ambitious rate. Using the Hasbrouck and Tindal (2004) norms for students in the 50th percentile, the whole group was expected to improve by an average of 4.5 wcpm during the same period. They surpassed this and improved by 14.63 wcpm, a rate that was 3.25 times the expected rate. Students continue to increase their oral reading rate after instruction, but improvement was slower than during instruction. The whole group increased by an average of 15.73 wcpm during the nine weeks between the posttest and the generalization test, a rate 1.94 times better than expected, though substantially less than the 3.25 times the expected rate seen during instruction. When comparing the actual oral reading rate changes over the entire fourteen-week duration of the study to the Hasbrouck and Tindal (2004) norms, the whole group and all subgroups continued to outperform their expected rate of progress. The average gain for the entire group was 30.35 wcpm, or 2.41 times the expected rate.

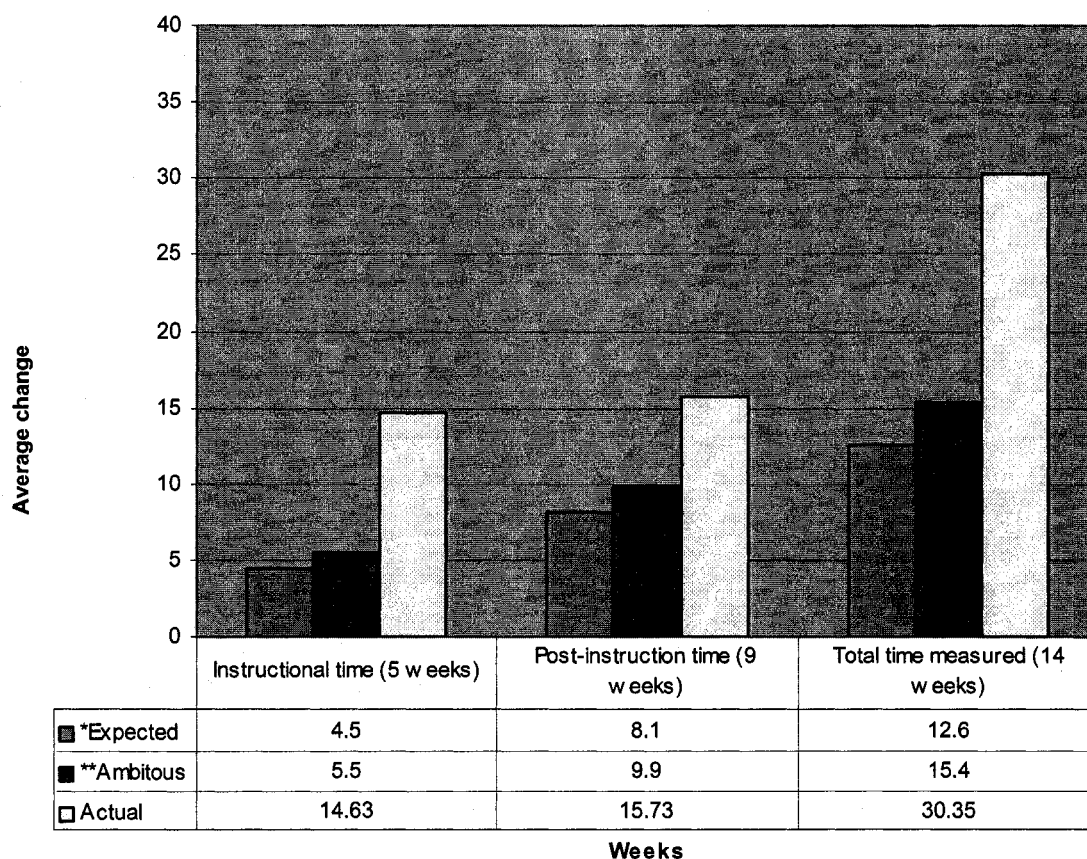


Figure 3: Oral reading rate increases for all students

In previous sections, it was explained that benchmark students showed the largest increases in their oral reading rate. Therefore, it is not surprising that they also made several times the expected or ambitious rates of growth illustrated in Figure 4. In the five weeks of instruction between pretest and posttest data, the ambitious goal for increasing rate was 5.5 wcpm. The benchmark students made the most impressive gains of all the groups in rate, averaging an additional 18.38 wcpm over the five weeks, or 3.34 times the ambitious rate. When the expected rate was used, the benchmark group's average improvement of 18.38 wcpm was 3.68 times better than expected. The students in the benchmark subgroup continued to make growth in reading rate after instruction, with an increase of 16.76 wcpm representing an improvement of 1.86 times their expected rate in nine weeks, but much slower than the 3.68 times the expected rate previously seen during instruction. Overall, when the total duration of the study is considered, the benchmark

students made the largest gains in words correct per minute, with average increases of 35.14 wcpm over the fourteen weeks, or 2.51 times the expected rate.

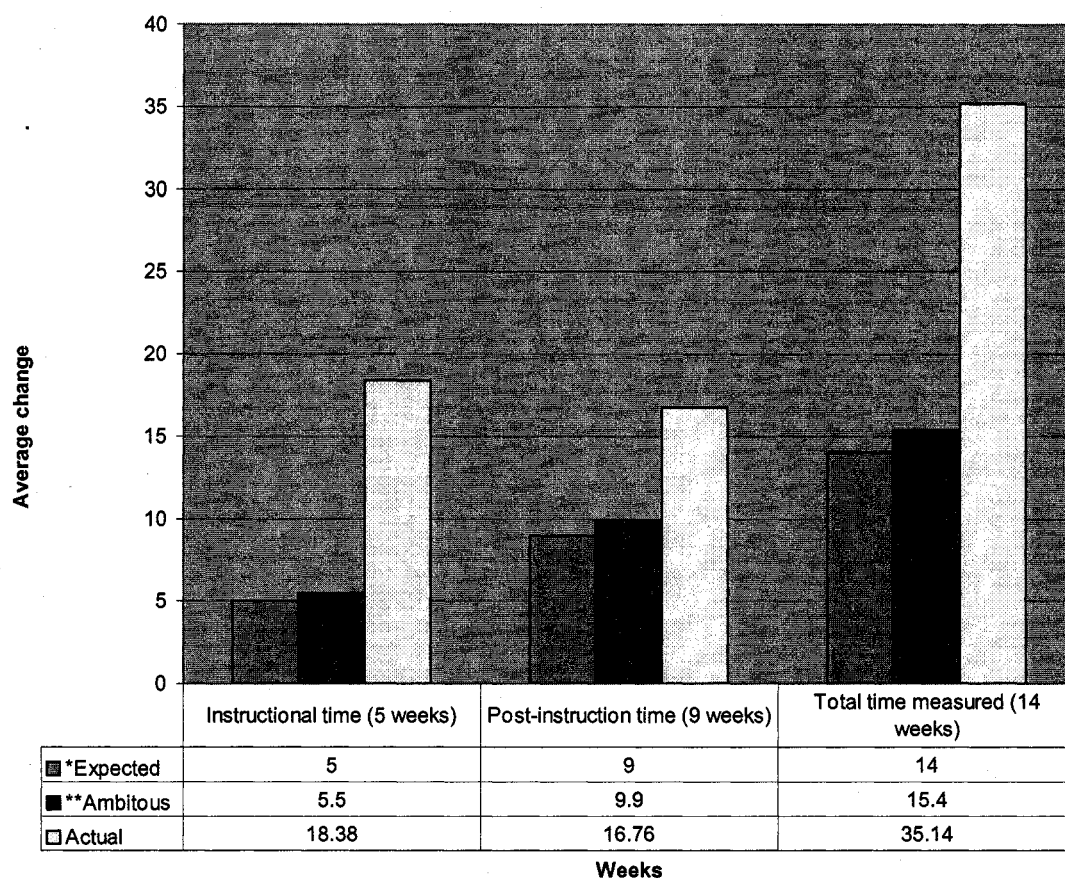


Figure 4: Oral reading rate increases for benchmark students

Even though the whole group and all subgroups made much better gains than either the “ambitious” or “expected” rates, it is especially noteworthy that the benchmark students, those who were on-track to meet or exceed spring grade level standards, made progress during instruction at a rate that was over three and a half times their expected rate of progress. This result addressed a key leadership question in this study. High performing students benefit from participating in whole class *REWARDS Intermediate* instruction. They are making some gains in accuracy and making substantial gains in oral reading

rate. Beyond the benefits to background knowledge and vocabulary yielded from simply reading more words, reading fluency has also been very highly correlated (.92) to comprehension (Snow et al., 1998). Practically speaking, teachers now have an answer for the student who says, "But I already know how to read these words." They can assure these students with confidence that they will be building their reading speed, which in turn will help increase comprehension while building vocabulary and background knowledge.

Strategic students, those in the mid-range, also made large increases in oral reading rates during this study (see Figure 5). In the five weeks of instruction between pretest and posttest data, the ambitious goal for increasing rate was 5.5 wcpm. Strategic students gained 14.42 wcpm, which is 2.61 times the ambitious rate. They also made outstanding growth in reading rate when compared to the Hasbrouck and Tindal (2004) expected rate values. They were projected to improve by 4.5 wcpm over the five weeks. Instead, they increased their rate by an average of 14.42 words, which was 3.2 times what was expected. This growth continued during the nine weeks after instruction ended. The expected rate of improvement during this period was slightly over eight words, so the strategic group's average improvement of 16.79 words is 2.07 times better than expected, though substantially less than their gains of 3.20. Finally, the strategic students improved by an average of 31.21 wcpm or 2.48 times the expected rate of improvement over the course of the entire study.

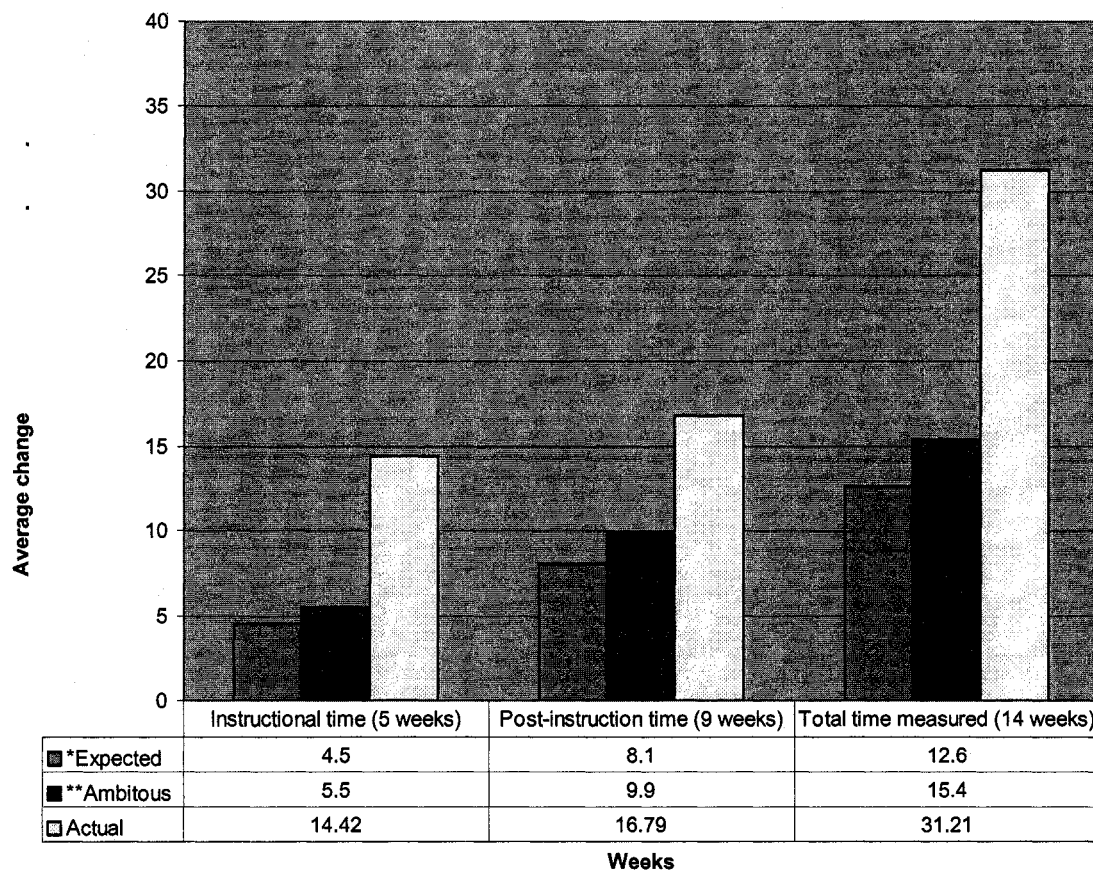


Figure 5: Oral reading rate increases for strategic students

Students who needed intensive intervention made improvements in their oral reading rate faster than anticipated when compared to ambitious or expected norms (see Figure 6). In the five weeks of instruction between pretest and posttest data, the ambitious goal for increasing rate was 5.5 wcpm. The results showed that intensive students gained 7.32 words, or 1.33 times the ambitious rate. When the five-week increases were compared against the expected gains for students in the 25th percentile, the intensive students' achievements appear to be even better. They were projected to add four wcpm over the five weeks of instruction, so their increase was 1.83 times the expected rate. Improvements continued after instruction ended. An increase of 12.64 wcpm during the nine weeks post-instruction was 1.76 times the expected value, which is similar to the 1.83 times achieved between the pretest and posttest. Finally, intensive students gained an average of 19.96 wcpm during the entire fourteen-week study duration. According to

the Hasbrouck and Tindal norms, they were expected to increase by 11.2 words. Therefore, their rate of improvement was 1.78 times the expected rate. While these gains are substantially smaller than those shown by the whole group and the other subgroups, it is important not to lose sight of the fact that the intensive students outperformed their ambitious and expected rates for oral reading rate increases in all phases of study. Furthermore, any disappointment in their rate of oral reading acquisition as compared to the other groups, is easily mitigated by their stellar improvements in accuracy as discussed previously.

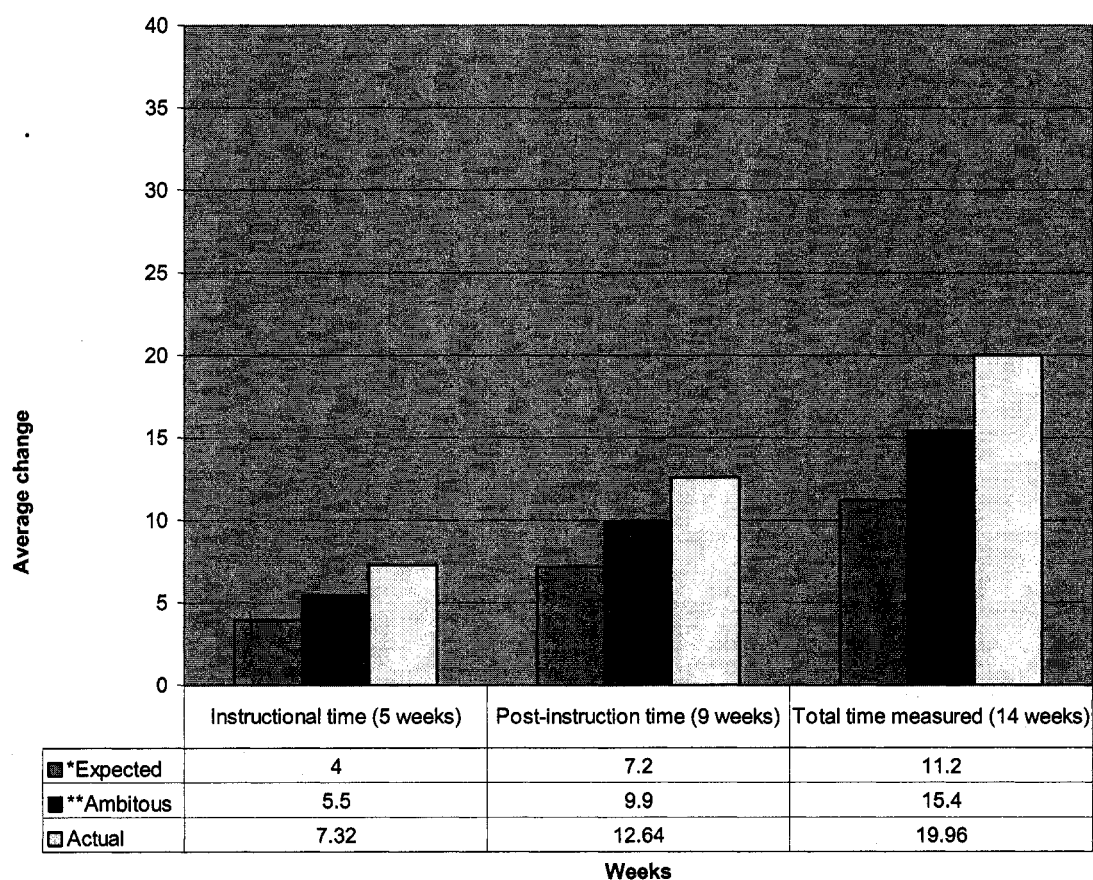


Figure 6: Oral reading rate increases for intensive students

Results summary

The whole group and all subgroups showed strong improvements in accuracy when reading multisyllabic words. Students in the intensive group made the largest gains both in taught and untaught words, followed by students in the strategic group, while the benchmark group, students who were already strong decoders, showed the least growth. Notably, the strategic subgroup of students showed the smallest decline in decoding accuracy when reassessed nine weeks after instruction. Even though all subgroups showed some decrease in accuracy over time, the changes were relatively small.

In contrast to the accuracy results, benchmark students made the largest gains in oral reading rate, while intensive students showed the least growth in rate. Gains for all subgroups continued in this area even after instruction. When study results were compared to published “ambitious” and “expected” rates of growth, the whole group and all subgroups substantially outperformed the anticipated rate of growth.

In short, these results seem to indicate that the *REWARDS Intermediate* reading program is effective when implemented in large group, heterogeneous fourth grade classrooms. Low performing students made substantial gains in accuracy, while high performing students dramatically increased their reading rate. Middle range students made large gains in both areas.

Study limitations

This study assessed changes in oral reading *rate*, not *fluency* (Hasbrouck & Tindal, 2004; Shaywitz, 2003; Snow et al., 1998). Fluency and rate are often used interchangeably. However, they are not the same. Rate refers to number of words read in a minute (wpm) or the number of words read correctly in a minute (wcpm). Fluency is a much broader term that encompasses rate as well as other factors including expression and phrasing. A person can read very accurately and at a very high rate of speed, but be completely unintelligible to an audience and rob the passage of any comprehension value. Good fluency requires accuracy, a reasonable rate of speed, and the expressiveness of the reader as a joint effort to enhance comprehension. In this study, rate in the form of words

correct per minute (wcpm) is used to increase the precision of the measurements and the reliability of the data collected. Furthermore, it is a standard format of measurement within the reading research community. Therefore, important features of fluency are sacrificed in lieu of increased precision in measurement and reliability of the data collected.

I also deliberately chose not to have a control group, or a class of students who did not receive the *REWARDS Intermediate* instruction. Given the direction of the district, I did not believe that I could justify to the school board or parents withholding instruction from a group of students from a promising intervention simply to provide a “control.” On the other side, I did not consider a control group from another district to be valid because of the far-reaching influence the philosophy of Levels of Intervention has already had on the Hoquiam School District. Buildings were in their second year of full implementation. A school in a different district without the district-wide training and emphasis will be at a disadvantage even before instruction begins. In short, by attempting to control for the instructional variable, I would have sacrificed the contextual variable.

Finally, this study addressed results based on the performance of individual students on a series of assessments. The study did not seek to address the reasons why these results were achieved. Understanding the inner mechanisms of advanced decoding skills may provide stable guidelines that support research-based strategies, which could be applied in a wider variety of situations.

One possible question which might result from this research project for future consideration is: What happens to students who were identified as intensive for reading difficulties when they were in kindergarten or first grade and given intensive reading instruction? Are these students more likely than other students to encounter difficulties at this level of advanced decoding too? Determining whether previously intensive students benefit significantly from intentional, explicit teaching and careful guidance through learning multisyllabic strategies may be an intriguing continuation of this topic for other researchers.

Conclusion and Implications for Leadership

In the beginning, I asserted that this project was not exclusively about the results of one isolated reading program, even though the previous results and discussion section illustrated the positive outcomes. Instead, this is another model of how leaders in schools or school districts can develop systems that routinely measure and readjust themselves to increase growth for all students as a whole and at the subgroup level. Both factual awareness and responsiveness are essential tools for leaders to use while ensuring that all students are growing and that the students who are behind make sufficient progress to meet grade level standards. This chapter highlights multiple implications for leaders including assessing the current conditions and contexts, marshalling resources, championing implementation, reviewing results, and modifying current practice.

Assessing the current conditions and contexts

The first area, assessing the current conditions and contexts, is accomplished through baseline assessment. This study used a combination of two assessments. *REWARDS Intermediate*, like many intervention programs, included a pretest. If the program does not have a pretest, it probably has a placement test. If neither exists, consider this to be a warning that the program may not be the type of research-based, explicit and systematic level of instruction needed to accelerate the progress of struggling learners. The program *may* be fine, but the lack of a way to monitor progress is enough to warrant a cautious review. In this case, the *REWARDS Intermediate* pretest provided three baseline measures. Two of the scores, word part accuracy and whole word accuracy, measured how well students were able to decode multisyllabic words in isolation. The third assessment, a one-minute oral reading fluency, provided a score of the students' facility with reading words in context.

Another assessment, the fourth grade fall *DIBELS* scores divided the students into three groups: intensive, strategic, and benchmark, according to the statistical likelihood that a particular student would fail to meet the spring benchmark. Most of the time, schools need to do additional diagnostic testing or a more in-depth error analysis using the *DIBELS* data in order to determine what the specific areas of need were for intensive and

strategic students. In the case of this study, the design was built around measuring the impact of whole group instruction on subgroups, so further analysis was unnecessary because we did not plan on using that information. This is a key concept that occasionally gets lost in the testing frenzy. If the data from testing are not going to be used to make instructional decisions, do not do it. Spend that time teaching students. Leaders need to help teams focus and articulate why each proposed assessment is needed and how the data will be used, which leads into the second leadership responsibility, marshalling resources.

Marshalling resources

The next area that has leadership implications is marshalling resources. There will always be a myriad of ways to spend the resources in a school or district. The leadership action here is to winnow away from the many good choices to focus one of the most important resources, identifying priorities, in a select area or two. To do this there were additional resources that needed to be gathered to support this type of review of an intervention. Specifically, this study needed to decide who was administering the assessments, who was teaching the intervention, who (and what budget) was purchasing the materials and who was collecting the data into a useable format that staff could understand in order to judge the effectiveness and efficiency of the intervention. Other districts or schools wanting to replicate this study would need to make similar decisions about resources. Hopefully by now, it is evident that resources are more than just money. In fact, the human resources needed to implement an intervention study are quite a bit more important than the money involved as we will see in the next section, championing implementation.

Championing implementation

With assessment complete and the resources available, the implementation of an intervention review might appear to be a straightforward process. Some people might presume that the next part is nearly automatic. This, of course, is not the case. It is incumbent upon leaders to continue to champion the implementation of the intervention review. Other leaders replicating this or a similar study should be prepared to operate as

the institutional memory. In other words, know the timelines for assessment and instruction, check in with the staff members teaching the intervention, follow-up and provide a forum for asking any questions that arise during the study. For example, mid-way through this study, one teacher wanted to divide each lesson and teach it over two days instead of one day. Her reasoning was that students would benefit from slower instruction with more practice; being exposed to the instruction for twice as many days would help the struggling learners; and that more class time would be available for other activities.

On the surface, the arguments sound valid. However, part of the leadership responsibility in championing implementation was knowing that for REWARDS Intermediate (and most direct instruction programs) a fast or “perky” pace is essential in order to maintain student interest and engagement. Arguably, slowing the instruction would hinder students’ learning more than helping it. The second point, that more days of instruction would benefit students who were struggling was at the heart of what this study was trying to determine. The third point was actually misleading math. Instead of spending 20-25 hours over 25 instructional days by completing one lesson a day, this teacher was proposing spending 38-42 hours over 50 instructional days. In spite of previous conversations and agreements about the structure and purpose of the study, this teacher needed to be reassured again that her students would get their needs met. Once we reestablished that the results would be reviewed carefully and changes would be made for the next group if it was discovered that struggling learners were not able to make good growth from the current pace and structure and that extra interventions would still be available to struggling learners after the study, she was able to support the study’s structure again. However, if there had been no one championing the implementation by responding to concerns and reestablishing agreements, it is likely that this teacher would have made a well-intentioned decision based on her professional judgment that would have eliminated her class as a source of standardized data and may have actually been detrimental to her students’ learning.

Reviewing results

The practice of reviewing results was the biggest change from past practice for the Hoquiam School District. Individual student performance was being carefully considered, but the effectiveness of an intervention on a whole group or subgroups was only considered occasionally and was based on intuition rather than data. To do this, a format had to be developed that would show changes in student performance during the five week instructional period between pretest and posttest; during the nine-week period after instruction; and the overall time frame including the fourteen weeks from start to finish. Since part of the purpose of this study was to use tools that were accessible to and easily understood by school and school district staff, simple differences and averages were calculated for the whole group and each of the subgroups.

Next, a decision had to be made about what was an acceptable level of growth. Research literature was able to provide norms for growth in oral reading rates and this was compared the progress of students in this study who used *REWARDS Intermediate*. Once again, the decision was made to simply express the relationship between actual outcomes and published norms as a function of multiples. A complete discussion of those norms and the results of the comparison can be found in Chapter 4.

Other schools or districts that were interested in conducting a similar review on an intervention program would need to decide the area(s) of growth being considered, the measurements needed (e.g. decoding accuracy, words correct per minute, comprehension accuracy), the time periods for comparison, the subgroup divisions, and assessments needed that would yield the type of measurements needed. Once that has been determined, the leader needs to allocate the financial and human resources and work with the instruction team to determine the personnel and timelines for each step of the investigation.

Modifying current practice

Finally, leaders must help staff use the results of data to change current practice. In the case of this study, the results seemed to indicate that the current practice of teaching *REWARDS Intermediate* in a heterogeneous, whole class format is effective and efficient for the whole group and the subgroups. As was explained in Chapter 4, effectiveness was judged by growth in accuracy and reading rate. The results of this study seemed to indicate that the whole group and all the subgroups improved their decoding accuracy and oral reading rate. It appears that the students who were at the lowest instructional levels at the start of the program improved the most in accuracy. Conversely, students who were already strong readers and were at the highest instructional levels before instruction made the greatest gains in oral reading rate as measured by words correct per minute. Efficiency, which asks if rate of growth was reasonable given the time and other resources allotted, was determined through comparison to established norms. Since all subgroups averaged well over the expected norms for rate of growth, and in some cases, made exceptional growth compared to the norms, the fourth grade teachers, administrators, and Levels of Intervention review team in the Hoquiam School District also judged this intervention to be an efficient use of time and resources. Therefore, even though they entered the process willing to make instructional changes based on data, Hoquiam will continue use *REWARDS Intermediate* as a heterogeneous, whole class intervention for all fourth grade students. One change they will be making, based on the experience of the fourth grade teachers, is to shift the instructional timeline from the beginning of the school year to mid-October. This gives the fourth graders time to learn the new routines of the class, to recoup their summer reading skill losses, and to experience the first unit of the core reading curriculum. The district is not sure that this will have any impact on students' ability to learn from the *REWARDS Intermediate* program, but they now feel confident in their ability to collect and analyze next year's intervention data, compare it to this year's data, and continue or modify their decision to teach the program later in the fall. This confidence and sense of group-efficacy is a tangible (and highly rewarding) demonstration of the second goal of this study for both the Hoquiam School District and the lead investigator.

Other leaders seeking to evaluate an intervention program need to build in a structure to use the analysis of results and foster a conversation about who is and is not being well-served by the intervention. When evaluating the results both effectiveness (is growth achieved and for whom) and efficiency (is the amount of growth reasonable based on student needs, as well as the resources invested in the intervention) need to be considered. From there, the leader needs to facilitate a process in which teachers and other instructional leaders brainstorm, evaluate, and select modifications to the current system designed to help underserved students. Once the decisions around modifying current practice are made, the leader must help the instructional team remember their decision over time and follow the timelines established by the group. The second part of leadership in modifying current practice is quite similar to the responsibilities in championing implementation. This makes logical sense because the group is now instituting a new instructional practice.

In the end, this study had a dual purpose. One purpose was to provide comprehensive and systematic data on the accuracy and oral reading rate changes for fourth graders when they receive heterogeneous, whole class instruction in *REWARDS Intermediate*. The results were compelling and illustrated that all students were learning, but that they were improving in different ways (reading rate v. decoding accuracy) By itself, this is valuable information to shape the practice of the Hoquiam School District and also to support the claims of effectiveness found in the *REWARDS Intermediate* teacher's guide. However, without the second purpose, the leadership implications, these results would be limited solely to a single, specific reading intervention.

Hopefully, the approach used in this study and the results that were obtained can also be applied on a much broader level as well. This study addressed a common current problem of practice that occurs when schools and districts get too focused on providing interventions based on assessment data and forget to evaluate the effectiveness and efficiency of those interventions, particularly on instructional subgroups. Ultimately, the leadership implications for action make this a social justice quest to provide all students,

including those in various instructional subgroups, with a strong literacy base from which to conduct and thrive throughout the rest of their lives.

References

- Adams, M. J. (1990). *Beginning to read: Thinking and learning about print*. Cambridge, MA: The MIT Press.
- Anderson, R. C., Hiebert, E., Scott, J. A., & Wilkinson, I. A. (1985). *Becoming a nation of readers: The report of the commission on reading*. Pittsburgh, PA: National Academy of Education.
- Archer, A. L., Gleason, M. M., & Vachon, V. L. (2005). *REWARDS reading excellence: Word attack & rate development strategies*. New York: Sopris West.
- Archer, A. L., Gleason, M. M., Vachon, V. L., & Hollenbeck, K. (2004). *Instructional strategies for teaching struggling fourth and fifth grade students to read long words*. Unpublished manuscript.
- Barr, R. D., & Parnett, W. H. (2003). *Saving our students, saving our schools: 50 proven strategies for revitalizing at-risk and low performing schools*. Glenview, IL: Pearson Professional Development.
- Bell, L. (2004). Strategies that close the gap. *Leadership Information*, 46-49.
- Borg, W., & Gall, M. (1983). *Educational research: An introduction* (Fourth Edition ed.). White Plains, NY: Longman, Inc.
- Borman, G. (2002). How can Title I improve achievement? *Educational Leadership*, 60(4), 62-68.
- Creswell, J. W. (2002). Collecting quantitative data. In *Educational research: Planning, conducting, and evaluating quantitative and qualitative research*. Upper Saddle River, NJ: Merrill-Prentice Hall.
- Cunningham, P. M. (1998). The multisyllabic word dilemma: Helping students build meaning, spell and read "big" words. *Reading and Writing Quarterly*, 14(2), 189-218.
- D'Angiulli, A., Siegel, L., & Maggi, S. (2004). Literacy instruction, SES, and word-reading achievement in English-language learners and children with English as a first language: A longitudinal study. *Learning Disabilities Research and Practice*, 19(4), 201-213.
- Davis, R. D. (1997). *The gift of dyslexia* (Second ed.): Perigee Books.
- Feifer, S. G., & De Fina, P. A. (1998). *The neuropsychology of reading disorders: Diagnosis and intervention workbook*.

- Fuchs, L. S., & Fuchs, D. (1993). Formative evaluation of academic progress: How much growth can we expect? *School Psychology Review*, 22(1), 27.
- Gerber, M., Jimenez, T., Leafstedt, J., Villaruz, J., Richards, C., & English, J. (2004). English reading effects of small-group intensive intervention for K-1 English learners. *Learning Disabilities Research and Practice*, 19(4), 239-251.
- Glicking, E. (2005). *Instructional assessment: Are you paying attention?* Paper presented at the NASP International Summer Conference 2005, Las Vegas, NV.
- Graves, A., Gersten, R., & Haager, D. (2004). Literacy instruction in multiple-language first grade classrooms: Linking student outcomes to observed instructional practice. *Learning Disabilities Research and Practice*, 19(4), 262-272.
- Haigler, K. O., Harlow, C. W., O'Connor, P. E., & Campbell, A. (1994). *Literacy behind prison walls: Profiles of the prison population from the* (No. 0160453828: \$13.00). Washington DC: U.S. Department of Education.
- Hall, S. L., & Moats, L. C. (1999). *Straight talk about reading: How parents can make a difference during the early years*: Contemporary Bks.
- Hasbrouck, J. (2004a). *Assessment & Washington's K-12 reading system*. Paper presented at the OSPI Summer Institute 2004, Yakima, WA.
- Hasbrouck, J. (2004b). *Washington's K-12 reading system*. Paper presented at the OSPI Summer Institute 2004, Yakima, WA.
- Hasbrouck, J., & Leppien, T. (2004). *SAILS: A K-12 reading system*. Paper presented at the OSPI January Conference 2004, Spokane, WA.
- Hasbrouck, J., & Tindal, G. (2004). Oral reading fluency: 90 years of measurement. *Behavioral Research and Teaching*, 33.
- Hintze, J. M., Ryan, A. L., & Stoner, G. (2003). Concurrent validity and diagnostic accuracy of the dynamic indicators of basic early literacy skills and the comprehensive test of phonological processing. *The School Psychology Review*, 32(4), 541-556.
- Langdon, T. (2004). DIBELS: A teacher-friendly basic literacy accountability tool for the primary classroom. *Teaching Exceptional Children*(Nov/Dec 2004), 54-58.
- Leafstedt, J., Richards, C., & Gerber, M. (2004). Effectiveness of explicit phonological-awareness instruction for at-risk English learners. *Learning Disabilities Research and Practice*, 19(4), 252-261.

- Lenz, B. K., & Hughes, C. A. (1990). A word identification strategy for adolescents with learning disabilities. *Journal of Learning Disabilities, 23*, 149-158+.
- McNamara, J. K., Scissons, M., & Dahleu, J. (2005). A longitudinal study of early identification markers for children at-risk for reading disabilities: The Matthew effect and the challenge of over-identification. *Reading Improvement, 42*(2), 80-97.
- McPeak, L. (2004). A vision of schooling in which all students achieve at high levels, *OSPI Summer Institute*. Yakima, WA.
- Miller, M. M., R. (1997). The importance of overt responding in programmed instruction even with added incentives for learning. *Journal of Behavioral Education, 7*(4), 497.
- Nagy, W. E., & Anderson, R. C. (1984). How many words are there in printed school English? *Reading Research Quarterly, 19*, 304-330.
- Perfetti, C. A. (1984a). Reading acquisition and beyond: Decoding includes cognition. *American Journal of Education, 93*, 40-60.
- Perfetti, C. A. (1984b). Some reflections on learning and not learning to read. *Remedial and Special Education, 5*, 34-38.
- Perfetti, C. A. (1986). Continuities in reading acquisition, reading skill, and reading disability. *Remedial and Special Education, 7*, 11-21.
- Perfetti, C. A., Beck, I. L., & Bell, L. C. (1987). Phonemic knowledge and learning to read are reciprocal: A longitudinal study of first grade children. *Merrill-Palmer Quarterly, 33*, 283-319.
- Rayner, K. (1985). The role of eye movements in learning to read and reading disability. *Remedial and Special Education, 6*, 53-60.
- Samuels, S. J. (1988). Decoding and automaticity: Helping poor readers become automatic at word recognition. *The Reading Teacher, 41*, 756-760.
- Shaywitz, S. (2003). *Overcoming dyslexia*. New York, NY: Knopf.
- Shepard, L. A., & Smith, M. L. (1990). Synthesis of research on grade retention. *Educational Leadership, 47*(8), 84-88.
- Slavin, R. E., & Madden, N. A. (1989). What works for students at risk: A research synthesis. *Educational Leadership, 46*(5), 4.

- Snow, C., Burns, M. S., & Griffin, P. (1998). *Preventing reading difficulties in young children*. Washington, DC: National Academy Press.
- Stanovich, K. E. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly, 21*, 360-407.
- Stringer, E. (2004). *Action research in education*. Columbus, OH: Pearson Education, Inc.
- Vachon, V. L., & Gleason, M. M. (2004). *The effects of mastery teaching and varying practice contexts on middle school students' acquisition of multisyllabic words reading strategies*. Unpublished manuscript.

VITA

Patricia Marie (Wheeldon) Zurybida, known as Tricia Zurybida, was born on May 13, 1970, the daughter of Josephine and Robert Wheeldon. After commencing from high school at Seattle Preparatory School in Seattle, Washington in 1987, she attended the Mateo Ricci College at Seattle University from 1987-1990 and graduated with a Bachelor of Arts degree. From 1990-1992, she attended the University of Washington in Seattle, where her emphasis was in education for students with severe disabilities. Tricia graduated with a Master of Education degree and teaching certificate in 1992. She taught in the Federal Way and Lake Washington School Districts from 1992-2002. She continued her education at the University of Washington in the Danforth Educational Leadership Program XIV and earned her Principal and Program Administration credentials in 2002. Beginning in the July of 2002, she has served as the Director of Special Services for the Hoquiam School District. Tricia returned once again to the University of Washington in the Leadership for Learning Program, Cohort 2 from 2004-2007. In the spring of 2007, she earned her Doctor of Education in Educational Leadership and Policy Studies with an area of emphasis in literacy. She lives in Hoquiam, Washington with her husband, Paul, her daughter Elizabeth, and her expected child, Baby Z, due in September 2007.