Meeting the Needs of Failing Readers:
Cautions and Considerations for State Policy

An Occasional Paper

by
Marsha Riddle Buly
Western Washington University

Sheila Valencia
University of Washington

April 2003
(Document O-03-1)

Center for the Study of Teaching and Policy
UNIVERSITY OF WASHINGTON

Center for the Study of Teaching and Policy

A National Research Consortium

UNIVERSITY OF WASHINGTON (lead institution)

STANFORD UNIVERSITY

TEACHERS COLLEGE/COLUMBIA UNIVERSITY

UNIVERSITY OF MICHIGAN

UNIVERSITY OF PENNSYLVANIA

Other active participants in CTP's research and dissemination program include researchers affiliated with Indiana University, Michigan State University, Pennsylvania State University, the University of California at Santa Barbara, the University of North Carolina, and Education Matters, Inc.

CTP studies the way policies and conditions in schools, districts, states, and teacher education institutions shape the quality of teaching and learning in the nation's elementary and secondary schools. The Center pays particular attention to the ways these policies and conditions interact with each other to influence the teaching profession and its practice.

The Center's program of research is carried out in collaboration with various other research organizations, among them other OERI-funded research centers, including the Consortium for Policy Research in Education (CPRE), the Center for Research on Education, Diversity, and Excellence (CREDE), and the Center on English Learning & Achievement (CELA). The Center is affiliated with a variety of professional and advocacy organizations that represent teachers, teacher educators, state and local policymakers, disciplinary groups, and educational reform interests.

Another version of this report appears under the title, "Below the Bar: Profiles of Students Who Fail State Reading Assessments" in Educational Evaluation and Policy Analysis, 24, 3, 219-239. We would like to thank Joe Jenkins, Mike Knapp, Scott Stage, Karen Wixson, Phil Buly, and anonymous reviewers for their helpful comments on an earlier draft of this manuscript.

Contents

Abstract

Every year thousands of students fail state reading tests and every year policymakers and educators search for strategies to help these students succeed. In this study, we probed beneath students' failing scores on a state reading assessment to investigate the specific reading abilities that may have contributed to student performance. We found that scores on state tests mask distinctive and multifaceted problems having to do with word identification, fluency, and meaning. Our findings are a caution to policymakers and educators who may be tempted to treat the same all students who score "below standard" on statewide reading assessments that now proliferate the education landscape. To do so is to miss the different instructional emphases called for by the underlying skills, strategies, and needs of failing students. Such a practice not only limits individual student progress; it may lead to an oversimplification of reform efforts and evaluation. This report presents reading profiles of failing students and discusses five areas—instruction; multiple indicators; alignment among standards, assessment, and instruction; allocation of resources; and evaluating reform—as potential policy levers for improving student performance in reading.

INTRODUCTION

In the field of reading, concerns about student performance and, indeed, student learning have prompted policymakers and school administrators to search for the "silver bullet," i.e., *the* program or instructional strategy that will improve student achievement. Anchoring the search is an assumption that what lies beneath students' poor performance on state reading assessments is a monolithic reading problem that can be solved with a similar—if not singular—fix (Allington, 2001; Allington & Walmsley, 1995; Duffy & Hoffman, 1999). Indeed, many schools districts and states use results of high-stakes assessments to make wholesale decisions about instructional approaches and related funding for the specific programs and the professional development they believe will help all the "not proficient" readers become "proficient". Yet often, the policymaking is crafted with little consideration of empirical evidence about the nature of the reading difficulties of failing students.

In this study, we provide empirical evidence that should help policymakers and educators understand better the complexity of reading performance and chart more deliberate courses to improve student learning and achievement. Heeding the caution about oversimplifying interpretations of test scores (Linn, 2000), standards-based reform (Cohen & Spillane, 1992; National Research Council, 1999; Valencia & Wixson, 2001), and the reading process itself (Lipson & Wixson, 1997; Snow, Burns, & Griffin, 1998), we looked beyond the scores on state reading assessments to examine the specific reading abilities that contribute to students' poor performance on state assessments. We also examined whether there are distinctive performance patterns that distinguish various groups of students.

Underlying this study is a long-standing body of research that suggests that reading involves a number of abilities or components including word identification, phonemic awareness, comprehension, reading fluency, and vocabulary (Adams, 1990; Chall, 1967; Clay, 1993a, 1993b; Gough, Hoover, & Peterson, 1996, National Reading Panel, 2000; Snow, Burns, & Griffin, 1998; Stahl & Murray, 1998; Stanovich, 1994). Other studies that bear on our investigation are those that raise concerns about the "influences" on policymakers and their thinking, including how they get information, the powerbrokers behind the scenes, and the actual studies and reports they rely on for policymaking (e.g. Fox, 2001; McDaniel & Miskel, 2001; McDaniel, Sims, & Miskel, 2001). Some of these studies suggest that policymakers may not have the depth of understanding they need to frame policy aimed at instructional reform, especially reform emphasizing ambitious pedagogy. As a result, they may miss or gloss over "the intended functions of the reforms focusing instead on the forms of the reform" (Spillane, 2000, p. 154). As policy becomes more specific and more focused on particular classroom instructional practices, policies based on misguided assumptions, misunderstandings, or skewed data sources become especially problematic.

In the field of reading, policymakers have begun to shift their attention to specific classroom practices and have initiated policies focused on two attention-worthy areas: redesigning classroom instruction for beginning readers (Allington, 2000; Snow, Burns, & Griffin, 1998) and implementing early interventions for struggling readers in the primary grades (Hiebert & Talyor, 1994). Many of these efforts at the state and local levels have targeted what is widely accepted as a critical component of early reading—decoding, and in particular, instruction in phonics (Paterson, 2000). Fueling this focus on beginning readers, interestingly, are the test results of intermediate grade students, the underlying assumption being that poor performance on, say, the National Assessment of Educational Progress

(NAEP) administered nationally to 4th graders reflects students' difficulties with more basic, earlier reading skills. Yet, since most high-stakes tests at the intermediate grade focus on comprehension, there is little data to suggest that working with early readers primarily on decoding will lead to increased reading performance of comprehension when those readers are tested in 4th grade.

The Study

This study took place in Washington state, a typical state in terms of its approach to school reform and one of 49 states that has implemented new standards and assessments in reading (Education Week, 2001). Washington operationalized a state system of assessment in 1997 called the Washington Assessment of Student Learning (WASL). The WASL is a criterion-referenced test aligned to state standards. In reading, the test assesses literal, interpretive, and analytic comprehension of both fiction and non-fiction text through multiple choice and open-ended items. Like most other state reading assessments and the NAEP, the WASL tests students at several grade levels: early intermediate (grade 4), middle school (grade 7), and high school (grade 10). In addition to the usual high stakes associated with statewide assessments, Washington State plans, in 2008, to begin using the results of the 10th grade WASL to grant or deny students a certificate of mastery that would be linked to a high school diploma. The plan is to indicate, on a high school diploma or certificate awarded with the diploma, the areas in which a student has demonstrated proficiency on the 10th grade WASL (e.g. reading, writing, math, and listening). Students who initially fail an area on the WASL will be given additional opportunities to retake individual sections. As a result, the WASL drives many decisions about instruction and about funding by the state legislature, the state Office of Superintendent of Public Instruction (OSPI) , and local school districts.

Following release of the 1998 results indicating that 43% of 4th grade students had not met proficiency on the reading assessment, the Washington State legislature allocated $9 million to K-2 classrooms to be used for instructional materials and professional development focused on research-based, beginning reading strategies. The bill stated that "reading instruction in the early grades must consist of a comprehensive program that builds upon the firm foundational skills of phonemic awareness, decoding (using sound / symbol relationships), and reading comprehension" (Washington State Legislature, 1998, p.1). Although in theory this statement acknowledges the complexity of skilled reading, many school districts interpreted this as a call for greater emphasis on phonics instruction. In fact, newspaper articles and influential politicians encouraged districts to adopt instructional materials and approaches that emphasized phonics (Ko, 1997). The Chairman of the Senate Education Committee wrote an editorial titled "Johnny can't read because Johnny needs phonics" in which he stated his belief that "READING IS PHONICS! WRITING IS PHONICS!" (Hochstatter, 1998). In effect, the policy as enacted was based on an assumption that students who had not demonstrated proficiency at grade four must not have acquired earlier fundamental reading skills, including phonics.

However, from the WASL score alone, derived from a group-administered measure that primarily captures comprehension, policymakers had little way of knowing about the specific reading capabilities or difficulties that may have contributed to students' poor performance. Lacking more detailed analyses of student performance, the only sure conclusion was that students who failed the test had not

adequately answered comprehension questions. Their problems with these questions might derive from weak decoding skills but might also very well reflect problems with other reading abilities such as vocabulary, fluency, or comprehension strategies.

And yet, many school districts placed an emphasis on phonics instruction. For example, one school district halted its Sustained Silent Reading (time when students are reading independently) in favor of more time for direct instruction of phonics. Another school district contracted with an outside vendor to provide intervention on speech sounds and relating sounds to letters for struggling students. Another placed all students who failed the 4th grade assessment in a special decoding program. And yet still another forbade teachers from using book rooms, with leveled books, in order to follow a scripted phonics program. These types of interventions, which mirror approaches in other states (Edmondson & Shannon, 2002; Goodnough, 2001), grow out of a concern for low student performance on the state assessment, yet they take a monolithic view of the problem.

METHOD

Participants

The 108 participants in this study came from 17 of 20 elementary schools within an ethnically diverse semi-urban school district of approximately 18,000 students. Fifty-seven percent of the students in the district were Caucasian and 43% were students of color (19% Asian or Pacific Islander; 11% African American; 11% Hispanic; 3% American Indian or Alaska Native). Approximately 11% of the students received special education services and 7% received "pull-out" English as a Second Language (ESL) services. Fourth graders in this district performed somewhat below the statewide performance average on the reading section of the WASL for three years, with 40% of them passing in 1997, 48% in 1998, and 52% in 1999. Because state guidelines allow few exemptions from the WASL, both second language learners and special education students are included in these reported percentages.

The students in this study had scored at Levels 1-2 out of 4 on the reading portion of the WASL, placing them below proficiency according to state standards. Those who had individualized educational plans (IEPs) for reading or who were receiving pull-out ESL instruction were excluded from the sample because in-depth assessments had already been conducted with them and because their individual needs were likely being addressed in supplemental programs. We did include the large number of failing students who came from homes in which English was not the first language yet who did not qualify or had been "exited" from ESL programs and therefore were not receiving special instruction.

Measures

To deconstruct students' poor performance on the WASL, we used multiple measures designed to assess abilities in word identification, phonemic awareness, comprehension, reading fluency, and vocabulary. We also compiled background data on students' writing ability, home language, and

socioeconomic status. Because several of the instruments we used measured more than one of the above components, we describe each instrument below and reference the components assessed. (see Table 1).

Table 1. Instruments and Components Measured

	Word ID	P.A.	Comp.	Rate	Express.	Vocab.
Woodcock Johnson-R • Letter-word ID • Word attack	X					
Qualitative Reading Inventory-II	X		X	X	X	
WASL 4th grade selection	X			X	X	
Peabody Picture Vocabulary-R						X
CTOPP • Phoneme Segmentation • Phoneme Deletion		X				

Woodcock-Johnson Psycho-Educational Battery-Revised (WJ-R).

We used two tests from the Woodcock-Johnson Psycho-Educational Battery-Revised (Woodcock & Johnson, 1990) to assess word identification ability: Letter-Word Identification and Word Attack, both individually administered. Word identification skill contributes to comprehension (Adams,1990; Gough et al.,1996; Perfetti, 1985). Often referred to generally as decoding or phonics, word identification refers to the ability to accurately identify words, both familiar and unfamiliar, and requires students to apply a variety of strategies including sound / symbol correspondences, letter patterns, structural analysis and sight recognition of high frequency words. Students who experience difficulty with word identification often struggle with comprehension because they simply cannot read the words or because they must devote a majority of their cognitive resources to identifying words, leaving few resources available for processing meaning.

The test of Letter-Word Identification requires students to read single and multisyllabic words, ordered in difficulty. The test of Word Attack assesses students' ability to apply sound / symbol relationships and structural analysis to pronounce unfamiliar single and multisyllabic words and pseudowords (e.g. fot, crimtop). This provides a measure of students' ability to apply decoding skills without relying on context clues or familiarity due to wide reading experience. Both tests provide normative scale scores and grade level equivalents. The test manual reports split-half reliabilities of .94 and .914 for the Letter-Word Identification and Word Attack subtests respectively. Overall, the Woodcock Johnson Reading tests correlate .83 with the Wide Range Achievement Test, a widely used measure of reading.

The Qualitative Reading Inventory II (QRI-II).

The QRI-II (Leslie & Caldwell, 1995) is an individually administered informal reading inventory in which students read both narrative and expository passages at various grade levels. It assesses students' reading level and specific reading abilities. As the student reads aloud, the examiner records each error

verbatim and later analyzes the errors quantitatively and qualitatively. Students also answer comprehension questions. Unlike standardized tests that require all students to read the same material regardless of their reading level, the QRI-II allows the examiner to administer a series of tasks to determine the level at which students can most effectively be instructed (instructional level). Therefore, students are not evaluated on material that is too easy or too difficult for them. Furthermore, because students read a range of passages, it is possible to examine how well they can comprehend when decoding is not a problem.

In addition to providing information on appropriate instructional levels, the QRI-II provides information on two types of reading accuracy (total accuracy and acceptable accuracy), reading comprehension, and reading fluency. Total accuracy is the percent of words read correctly when reading leveled passages. Acceptable accuracy is the percent of words read correctly when only uncorrected, meaning-change errors are counted (i.e. a student's reading of "house" for "home" is not counted as an error; "house" for "horse" is counted as an error). Thus, acceptable accuracy provided a way to analyze the quality of the errors, not simply the quantity of errors, The QRI-II also provides a measure of comprehension by having students orally answer comprehension questions. This eliminates the confounding effects, found in many performance-based assessments, of using written responses to judge reading comprehension. Finally, the QRI-II provides a measure of one aspect of reading fluency, rate or speed of reading connected text (measured in words per minute), which has been shown to be correlated with comprehension (Adams, 1990; Carver, 1990; Pinnell et al., 1995). The QRI-II test manual reports interrater reliabilities of .99-.98 for accuracy and comprehension scores and concurrent validity correlations of .90-.75 with the Woodcock Reading Mastery Test-Revised.

To obtain a measure of reading expression, a second aspect of reading fluency, we used a holistic 4-point scale developed as part of a special study of the 4th grade National Assessment of Educational Progress (NAEP) as we listened to students read leveled reading passages from the QRI-II (Pinnell et al., 1995). The scale includes attention to phrasing, adherence to the author's syntax, and expressiveness. Although highly correlated with rate, these aspects of fluency, which we label "expression," are different in that they reflect attention to meaning. Research studies have shown these aspects of expressive fluency to be significantly related to reading comprehension (Dowhower, 1987; Pinnell et al., 1995).

To more accurately represent students' levels of performance, we transformed the QRI-II instructional grade level scores to reflect both grade level and performance level.[1] For example, a student who scored 80% comprehension on a 4th grade instructional level received a transformed score of 3.2 (4.0 x .80) while a student who scored 90% on a 2nd grade instructional level received a score of 1.8 (2.0 x .90). This transformation was calculated for all total accuracy, acceptable accuracy, and comprehension scores derived from the QRI-II.

Washington Assessment of Student Learning (WASL).

We supplemented the standard scoring of the WASL with students' oral reading of one narrative and one expository reading selection from the 4th-grade test (each approximately 200 words). This provided a direct measure of students' ability to decode the actual test selections. Scores for total accuracy, acceptable accuracy, rate, and expression were calculated using the same procedures as described above for the QRI-II.

Like most other state performance-based reading assessments, comprehension on the WASL is assessed using a mix of multiple-choice and open-ended questions. Alpha coefficients are .79 for both item types and .87 for the total test; interrater reliability for open-ended items is .98. Concurrent validity with the Comprehensive Test of Basic Skills-4 Reading Total is reported as .74 (Taylor, 1998). Scale scores on the WASL range from 0-700; the cut score for proficiency (Level 2) is 400 and the cut score for Level 1, the lowest level, is 375.

Peabody Picture Vocabulary Test-Revised (PPVT-R).

Vocabulary knowledge has long been recognized as a significant factor in reading comprehension (Anderson & Freebody, 1981; Davis, 1968; Nagy, 1988). To obtain information on receptive vocabulary knowledge, we administered the Peabody Picture Vocabulary Test-Revised (PPVT-R) (Dunn & Dunn, 1981) to each student. In this test, a student looks at four pictures while the examiner asks the student to point to one by name, thereby testing receptive vocabulary independent of the student's ability to decode words. The PPVT-R provides a normative scaled score and age equivalent (mean = 100, SD = 15). The manual reports test retest reliability of .77 for standard scores.

Comprehensive Test of Phonological Processing (CTOPP).

Research indicates a positive relationship between phonemic awareness (the ability to recognize, identify, and manipulate the smallest units of sounds) and reading proficiency. Recently, much attention has been focused on the potential effectiveness of phonemic awareness interventions for struggling beginning readers (Bus & Van Izendoorn, 1999; National Reading Panel, 2000; Snow et al., 1998; Wagner, Torgeson, & Rashette, 1994). We administered the phoneme deletion and phoneme segmentation subtests from the CTOPP (Wagner, Torgesen, & Rashette, 1999), the two highest levels of phonemic awareness, and therefore most appropriate for intermediate grade students (Adams, Foorman, Lundberg, & Beeler, 1998; Hoien, Lundberg, Stanovich, & Bjaalid, 1995). In the deletion task, students repeat a word the examiner says deleting either the initial or ending consonant. In the segmentation task, students are asked to repeat a word and then to say the word one sound at a time without any prompts.

We used a prepublication version of this normed measure. The mean standard scores for 10-year-old students, the closest age to the students in this study at the time of testing, were reported as 20.4 for the deletion task and 15.87 for the segmentation task. (J.R. Torgeson, personal communication, October 1998). Average alpha coefficients are .95 for the deletion task and .90 for the segmentation task. Test-retest reliability is reported as .79.

Data Collection

Seven examiners administered and scored the assessments. All were certified teachers; five had advanced degrees in literacy, the other two were recognized for their in-depth knowledge of struggling readers. Each examiner was trained to administer only specific measures, thereby increasing her expertise and reliability. For administration of the WJ-R and PPVT-R, examiners were trained and then administered the first test in teams to standardize procedures. For the QRI-II and the oral reading portion of the WASL, interrater agreement was established using tape recordings of students reading. Once acceptable agreement had been established (95% or better on overall scores and instructional levels), examiners individually assessed students with interrater agreement calculated on 10% of the cases over time. Interrater agreement ranged from .87 -.99 across measures. The order of administration of all the measures was counterbalanced.

The state administration of the 4th grade WASL occurs each April, with proficiency scores available by the end of August. The individual measures for this study were administered in September and October, shortly after WASL proficiency scores were made available and shortly after the students began 5th grade. Testing took place over a two-week period within each school, requiring a total of six weeks for testing of the entire sample. Examiners moved from school to school working with students, all of them in 5th grade, on specific measures. In each school, examiners were introduced to the students by their classroom teachers and were told that the purpose of this project was to learn more about students' reading. Students could opt out of testing or reschedule if they preferred not to miss a classroom activity; none declined to participate. During testing, students worked individually with examiners outside the classroom in a quiet place in the school. Examiners first spent time establishing rapport with the students and acquainting them with the testing process. Students were instructed to read as they normally would in the classroom or on their own and were assured that their performance would have no influence on their class grades.

Results

Descriptive statistics are presented in Table 2. In general, the results suggest that students who fell below proficiency on the 4th grade WASL reading assessment demonstrated skills below their current grade placement on all the measures: word identification, phonemic awareness, comprehension, vocabulary, rate, and expression. The WASL scale score correlated significantly with all the measures except the two phonemic awareness measures. Although phonemic awareness is implicated in studies of early reading, other studies have found that for more advanced students, spelling ability may interfere with assessments of phonemic awareness (Adams, 1990; Ehri, 1992; Scarborough, Ehri, Olson, & Fowler, 1998). In this sample, that seemed to be the case for many students. Qualitative analysis of student errors on CTOPP tasks revealed that, even after several practice items, many students identified sounds corresponding to the correct spelling of stimulus words rather than simply identifying sounds that could be heard (e.g. for the word beast, providing sounds for b/e/a/s/t/ rather than b/e/s/t/). Because student performance on the two measures of phonemic awareness appeared to be confounded by spelling ability and because these measures were not significantly correlated with the WASL score, they were not used in the subsequent factor analysis and cluster analysis.

Table 2. Descriptive Statistics

Variable	Mean	SD
Word Identification		
QRI-II Expos. Acceptable Accuracy *	3.1	1.1
QRI-II Expos. Total Accuracy *	3.1	1.1
QRI-II Narr. Acceptable Accuracy *	3.3	1.1
QRI-II Narr. Total Accuracy *	3.3	1.1
WASL Expos. Acceptable Accuracy **	93.0	6.7
WASL Expos. Total Accuracy **	89.0	8.3
WASL Narr. Acceptable Accuracy **	94.8	5.3
WASL Narr. Total Accuracy **	90.5	6.8
WJ-R Letter-Word Identification ***	92	15
WJ-R Word Attack ***	90	18
Phonemics Awareness (raw score)		
CTOPP—Segmentation	13.3	4.6
CTOPP—Deletion	17.0	4.5
Comprehension (grade equivalent)		
QRI-II Comprehension on Expos.	2.6	1.3
QRI-II Comprehension on Narr.	2.8	1.0
Vocabulary (standard score)		
PPVT-R Vocabulary	90	21
Rate (words per minute)		
QRI-II Rate Narr.	90.6	28.0
QRI-II Rate Expos.	88.5	29.6
WASL Rate Narr.	81.0	32.6
WASL Rate Expos.	78.8	34.1
Expression (1-4 rubric)		
QRI-II Fluency Narr.	2.5	0.9
QRI-II Fluency Expos.	2.5	0.9
WASL Fluency Narr.	2.2	0.9
WASL Fluency Expos.	2.3	1.0

Notes:
* Grade equivalent
** % correct
*** standard score

Factor Analysis

We conducted exploratory factor analysis to reduce the data and to identify variables with common underlying constructs. Because we expected all the factors to relate, the factor solution was obliquely rotated. We only considered variables with a loading of .7 or above in each factor. Three clear factors emerged, accounting for 78% of the variance on the WASL scale scores. We labeled these factors: word identification, fluency, and meaning (see Table 3), and discuss them one by one in the next section.

Table 3. Factor Analysis

Variable	Word Identification	Fluency	Meaning
QR-II Total Accuracy, Nar.	**.936**	-.561	.371
WASL Acceptable Accuracy, Nar.	**.932**	-.575	.283
QR-II Acceptable Accuracy, Nar.	**.924**	-.566	.384
WASL Acceptable Accuracy, Exp.	**.923**	-.551	.305
WASL Total Accuracy, Nar.	**.916**	-.640	.161
WASL Total Accuracy, Exp.	**.913**	-.643	.194
QR-II Total Accuracy, Exp.	**.908**	-.528	.486
QR-II Acceptable Accuracy, Exp.	**.901**	-.503	.523
WJ-R Letter-Word Identification	**.852**	-.736	.160
WJ-R Word Attack	**.762**	-.632	-.060
WASL Rate, Nar.	.616	**-.901**	.010
WASL Rate, Exp.	.596	**-.879**	-.005
QRI-II Rate, Nar.	.453	**-.877**	-.119
WASL Fluency, Nar.	.660	**-.870**	.368
WASL Fluency, Exp.	.699	**-.866**	.296
QRI-II Rate, Exp.	.453	**-.844**	-.201
QRI-II Fluency, Exp.	.558	**-.800**	.197
QRI-II Fluency, Nar.	.521	**-.759**	.243
QRI-II Comprehension, Exp.	.584	-.269	**.775**
PPVT-R Vocabulary	.205	-.061	**.761**
QRI-II Comprehension, Nar.	.591	-.329	**.706**

* Bold variables are those used for each factor

FACTOR #1: Word identification

All the measures of decoding (i.e. real word and pseudoword reading in lists and word reading in context) comprised this factor. The average standard score for students when reading pseudowords (WJ-R Word Attack) was 93, corresponding to a grade equivalent of 4.2. When identifying real words in a list on the WJ -R Letter Word Identification test, students' grade equivalent scores averaged 4.7 (standard score = 93). Although scores were somewhat lower than students' grade placement in school, their performance indicates skill beyond a basic, or beginning, level of sound / symbol correspondence.

When reading words in context from the QRI-II reading passages, the students' transformed scores averaged beginning 3rd grade level of word identification. However, more than half the students were able to accurately read enough words on the 4th grade QRI passages to place them at an instructional level of 4th grade or higher. Furthermore, when reading the actual WASL passages (passages that the test publisher determined were at 4th grade level), more than 60% of the students could correctly read enough words in these passages (>90%) to place them at instructional level. This suggests that either the reading level of the WASL passages may be a bit easier than a 4th grade instructional level or that many of the students may have word identification skills close to, or better than, a 4th grade instructional level.

Together, the direct measure of word reading on the WASL, performance on the 4th grade QRI selections, and WJ-R scores suggest that for many students, poor performance on the test was likely *not* due to a fundamental lack of decoding or word identification ability.

FACTOR #2: Meaning (Comprehension and Vocabulary)

The comprehension and vocabulary measures (QRI-II narrative comprehension, QRI-II expository comprehension, PPVT-R) loaded on this factor. It is important to note that all the measures required students to provide oral responses, eliminating the potential problem of confounding reading comprehension with writing ability, a potential problem with many comprehension tests, such as the state assessment, that use open-ended items (Goldberg & Kapinus, 1993, Jenkins, Johnson, & Hileman, 2001). However, even with the oral mode of response, students struggled to demonstrate meaning on the two comprehension measures from the QRI-II and the vocabulary measure from the PPVT-R. Grade equivalent scores on the QRI-II comprehension measures were high second grade; performance on the PPVT-R was a scaled score of 90, almost 2/3 standard deviation below the mean normative score of 100.

Although difficulty with word identification could contribute to difficulty with comprehension, overall these students could identify words at a higher level than they could comprehend. For example, only 24% of the students scored at a 4th grade instructional level for comprehension on the narrative QRI passage, yet 65% of the students were able to adequately decode the words in these passages at an instructional level (i.e. 90% accuracy or better). There were similar discrepancies between decoding and comprehension for the expository passages. These data suggest that many students struggled more with meaning than they did with word identification.

FACTOR #3: Fluency (Rate and Expression)

The measures of reading rate and of expression comprised the third factor. The importance of fluency is supported by a body of research that suggests that readers must be accurate when they read and they must also read quickly and with expression (LaBerge & Samuels, 1974; National Institute of Child Health and Human Development, 2000; Spear-Swerling & Sternberg, 1994). If decoding is slow or if readers are constantly stopping to figure out unknown words, cognitive resources for comprehension become limited and understanding may suffer.

Unlike the measures of word identification and comprehension, which provide a standard score or grade level reference point, grade level performance is more difficult to identify for rate and expression.[2] Pinell et al., (1995) used measurement procedures similar to those used in this study to examine the reading rates of a nationally representative sample of 4th grade students who participated in the NAEP oral reading study. The rate for 4th grade students who fell within the "basic" category, the lowest categorization on NAEP denoting "partial mastery of prerequisite knowledge and skills that are fundamental for proficient work" (National Center for Education Statistics,1996, p. xi) was generally between 104 and 129 words per minute. Students who scored below the "basic" category on the NAEP reading test read with an average rate of 104 words per minute or less. In contrast, students in this study averaged just 85 words per minute, substantially below both the "basic" and "below basic" students on NAEP.

In the area of expression, students in this study scored, on average, just over 2 on a scale of 1-4. Students in the NAEP sample who scored 1–2 on the same scale were judged to be nonfluent readers; on average they read at a rate of 89 words per minute or less and they scored below the "basic" level on comprehension. Students who were scored 3-4 on the NAEP scale (labeled "fluent readers") had an average reading rate of at least 126 words per minute. Overall, students in this study were experiencing significant difficulty in fluency—both rate and expression.

Cluster Analysis

The mean descriptive data and factor analysis suggest a particular pattern of student performance—one in which students' performance is slightly below grade level in word identification but substantially weaker in comprehension and fluency. These scores are based on group averages rather than on scores or patterns of individual students within the larger group. However, the results of a cluster analysis revealed that the pattern suggested by the averages for the whole group does not fit all, or even the majority, of the students. Rather, 10 distinct patterns were suggested in the cluster analysis.

Cluster analysis is a statistical procedure that initially combines the most similar pair of cases and then adds each new case in the sample to the best-fitting cluster until all cases are placed (Aldenderfer & Blashfield, 1984; Everitt, 1974; Romesburg, 1984; Speece, 1994).[3] In this analysis, each student represents a case and each case consists of three factor scores (word identification, meaning, fluency), which were derived by averaging a student's standardized variable scores for each factor. Our intent was to

Table 4. Cluster Analysis

Cluster	N	Word ID	Meaning	Fluency	WASL Score	% Not Eng.	% Low SES	Writing Score	% L1	% L2
1	10	.37	-.58	.79	374	60	90	49	50	50
2	9	.64	-.33	.48	378	67	89	47	44	56
3	16	-.07	-.34	.72	380	56	81	53	31	69
4	19	-.42	.49	-.27	384	16	42	49	32	68
5	11	.09	.75	-.40	388	09	36	58	09	90
6	15	.37	.48	-.31	380	27	67	60	47	53
7	12	.38	-.34	-.09	377	58	67	57	50	50
8	6	.34	-.13	-.55	372	50	67	46	50	50
9	2	-.255	-1.68	-.92	359	50	100	25	100	0
10	8	-.107	-.31	-.44	374	12	75	38	63	37

Key

- WASL Score = scale score (400 = proficient)
- Not Eng. = % students with home language reported to be other than English
- Low SES = % students eligible for free or reduced-price lunch
- Writing = % of possible points on content, organization, and idea section of the state writing assessment (64% = state average for all students)
- % L1 = Percentage of students who scored in Level 1 (lowest level) on the reading section of the WASL
- % L2 = Percentage of students who scored in Level 2 on the reading section of the WASL

determine whether the pattern suggested by the overall descriptive data fit a majority of the students or, alternatively, whether there were different compelling patterns of student performance.

Table 4 displays factor profiles for the students in each of the 10 clusters and pertinent background information for each (e.g. WASL score and level, home language, socioeconomic status, writing proficiency). In addition to empirical support for the derivation of the cluster solutions, the clusters also portray interpretable profiles from an educational perspective. From the average scores, we might conclude that students had similar needs, however as the cluster analysis reveals and, as we elaborate below, they are quite different. To make this point, we describe prototypical students from the clusters.

Clusters 1 & 2: Automatic Word Callers (18%)

The students in Clusters 1 and 2 are stronger in word identification and fluency than in meaning. We call these students "automatic word callers" because they can read the words quickly and accurately but fail to read for meaning. More than 60% of automatic word callers are second language learners; most of them are poor and receive free and reduced price lunch. While both Tim's and Marja's profiles reflect the pattern of automatic word callers, their reading strengths and needs differ from each other in subtle ways as exemplified in the following examples. Thus they, and other students like them, fit into different clusters.

Tim was a native English speaker, and he was poor. He was very good at identifying words and applying phonics to nonsense words, with scores averaging 9th grade on the Woodcock-Johnson Letter-Word Identification and Word Attack tests. On the narrative passage from the state assessment, his total accuracy was 98%, an indication of his ability to independently and easily read the 4th grade passage. His reading rate was among the highest of the students in this study—a rate of 174 words per minute. Although his standard score on the receptive vocabulary measure placed him in the average range (105), his comprehension scores on the QRI-II were low. He scored at a second grade level on the narrative passage and at a third grade level on the expository passage.

Although Tim's reading comprehension was low, he could produce written responses to prompts and express his ideas in writing. When compared with all students across the state, Tim scored in the top level of proficiency on the state writing assessment in content, organization, ideas, and conventions. We concluded that Tim's writing ability probably did not contribute to his low performance on the state assessment. However, when we analyzed Tim's reading performance on the state assessment, he scored only 5% of the total points possible on open-ended response items and 48% of the total points possible on multiple-choice items. So rather than having a basic problem in writing, Tim appears to have difficulty with comprehension; he struggles to generate written responses and answer questions related to what he has read.

Tim appears to have developed necessary skills in word identification, fluency, writing, and vocabulary that should enable him to easily comprehend and demonstrate his understanding. Yet, he struggles. From the data we have, it is impossible to determine if he is simply reading too fast and not attending to meaning or if comprehension is actually his primary problem. Before determining appropriate instruction for Tim, his teacher would need to have him slow his reading pace and then see if comprehension improved.

In contrast to Tim, some students in this cluster, like Marja, demonstrated difficulty in word meaning (PPVT-R). Many of these students were second language learners who were still developing vocabulary in their second language. When learning a second language, it is relatively easy to learn the surface language, the language that allows one to communicate with others (Cummins, 1991). However, it can take several years for an ESL student to learn the complexities of a second language, and to develop a degree of cognitive ability and vocabulary in the second language that might be needed to read and understand new material (Collier, 1987, 1989). The majority of second language learners, however, frequently only receive language support for a limited amount of time. As mentioned earlier, the students in this study had already been exited from ESL classes and were expected to succeed in mainstream classrooms. Classroom teachers, many of whom did not have ESL experience or preparation, were expected to meet their needs. Marja, like many other ESL students, would benefit most from opportunities to build her English vocabulary and comprehension of longer texts through both reading and oral language development.

Marja, Tim, and automatic word callers are relatively strong in word identification and fluency yet they struggle with meaning. Nevertheless, depending on their unique comprehension and vocabulary abilities, they would likely benefit from different instructional approaches and emphases.

Cluster 3: Struggling Word Callers (15%)

Students in Cluster 3 look somewhat similar to those in Clusters 1 & 2; they are relatively stronger in fluency and word identification skills than they are in meaning. However, unlike the students in Clusters 1 and 2, students in Cluster 3 are experiencing some difficulty in word identification.

Dang is a Struggling Word Caller. His word attack and word identification skills were uneven; he appeared to perform better when reading lists of words, as on the Woodcock-Johnson tests where he averaged 5th grade equivalency. However, when reading longer, more complex passages where he was also expected to comprehend, his word identification suffered. Initially, he scored at the frustration level (84%) for word identification when reading a 2nd grade QRI-II passage although he was able to self-correct while reading to score 98% acceptable accuracy. Even with this level of self-correction, he struggled to comprehend, scoring only 66% of the comprehension points on the 2nd grade passages, again a frustration level. In addition to decoding difficulty, vocabulary likely contributes to Dang's low comprehension score. With Vietnamese as his home language, Dang's score on the Peabody Picture Vocabulary Test was only 48—several standard deviations below both the norm group and below the students in this sample. Clearly, he is struggling with word meanings as well as comprehension of longer text.

Remarkably, reading at a 2nd grade level with a significant number of self-corrections, Dang read at a rate of 122 words per minute. It seems that Dang believes that good reading is fast reading. This rate may, in fact, be too fast for a student who is struggling with both decoding and with comprehension. For Dang, instruction in decoding, oral language, and comprehension, all in a rich meaningful context, seem warranted.

Cluster 4: Word Stumblers (18%)

Sandy's cluster comprises students primarily from English speaking families, the majority of whom are not considered poor, especially compared to students in the other clusters. Meaning is a relative strength for Sandy; she has considerably more difficulty with word identification.

Sandy's scores on the Woodcock Johnson tests indicate significant difficulty in the area of decoding; on the Word Identification test, she scored at the 2.8 grade equivalency and on Word Attack, she scored at the 1.9 grade equivalency. This is a student at the very beginning stages of word identification. The other measures of word identification also substantiate decoding problems. For example, Sandy's oral reading of the 4th grade WASL passages and QRI-II passages revealed decoding total accuracy at approximately 82%, placing her at a very low, frustration level even when she was reading 3rd grade QRI-II passages. This means that Sandy stumbled on so many words initially that she could not be expected to successfully read and comprehend. Surprisingly, however, Sandy persisted as she tried to decode, working until she could figure out some part of each word and then using context clues to help her get the whole word. In fact, Sandy over-relied on context because her word identification skills were so weak. As a result of her persistence and reliance on context, Sandy was able to correct many of her initial errors, which raised her acceptable accuracy scores to an instructional level. Remarkably, she was able to move from an 80% accuracy level on the 4th grade state selection to 99% acceptability when her self-corrections were taken into account. This is a student who strives for understanding.

As might be expected, Sandy was a slow reader. She spent a good deal of time trying to read the words and then trying to use the context when word attack failed her. Her rate averaged just 77 words per minute. In general, she read haltingly, word by word. Once she identified the words, however, she seemed to know their meanings and to get the overall meaning of the passages she read. This is remarkable given the number of words she struggled to read. It is likely that Sandy's comprehension was aided by her strong language background and receptive vocabulary, which was more than two standard deviations above average for her age.

The evidence suggests that Sandy has considerable difficulty with word identification even at the basic level of sound / symbol correspondence and high frequency words. It is also likely that her word identification difficulties contributed to her low fluency scores. She understands that reading should make sense and relies on that to compensate for poor word identification abilities. Clearly, instruction targeted at word identification and fluency at her instructional level would be beneficial for Sandy, along with continued emphasis on reading for meaning.

Cluster 5 & 6: Slow and Steady Comprehenders (24%)

Both Steven and Stacee were members of this cluster; they came from homes in which English was spoken, performed in the average range on the Peabody Picture Vocabulary Test, and scored at a proficient level on the state writing assessment. Like the other students in these clusters, Steve and Stacee read slowly, yet their word identification and comprehension abilities were relatively strong. Although they have the same general profile, they also exemplify interesting differences that have implications for instruction.

Steven's average word identification was at the high 3rd grade level, below grade level but not dramatically so. As indicated on the Woodcock Johnson tests, he was stronger at identifying real words

than he was at applying decoding strategies to pseduowords or multisyllabic words. In fact, on both the 4th grade WASL selections and the 4th grade QRI-II selections, Steven was able to read the words quite adequately, but his total accuracy was lower than his acceptable accuracy. In other words, Steven had some initial difficulty with decoding but he was able to self-correct many of his errors and eventually read 4th grade selections at his instructional level. Decoding difficulties at this grade level generally do not suggest that students need basic instruction in phonics, as in Sandy's case (Cluster 4), but more likely indicate some difficulty with advanced decoding skills such as reading multisyllabic words.

Steven was persistent, striving to read accurately and to read for meaning. However, his rate ranged from 34 to 61 words per minute across both 4th grade QRI-II selections and both WASL selections, a seriously slow pace. His tendency to self-correct, while a sign that Steven was reading for meaning, contributed to this slow rate of reading. Nevertheless, Steven comprehended what he read, scoring at a 4th grade instructional level, the highest level at which he was tested. His age appropriate performance on the Peabody Picture Vocabulary Test also supports his relative strength in meaning.

Stacee was stronger than Steven in the area of word identification, scoring at the 7th grade equivalency on both the Woodcock Johnson tests and above instructional level for word identification on both the 4th grade WASL and QRI-II passages. Like Steven, she was a good comprehender and, like Steven, she was slow. Her rate ranged from 56 to 69 words per minute, substantially below both her grade level and others who struggled with the state reading assessment.

The outstanding characteristic of both Steven's and Stacee's profile is their extremely slow reading rate coupled with strong comprehension. Although research would suggest that such a slow rate might interfere with comprehension, neither student suffered comprehension difficulties. The fact that the WASL is an untimed test likely helped both students who were within several point of passing. However, this close analysis of their performance suggests that even if they had passed the test, both students would still have considerable difficulty in rate and Steven would have some difficulty decoding multisyllabic words, two problems which would likely impede future reading success. Furthermore, with such a lack of automatic word identification skills, it is unlikely that either student enjoys reading or spends much time reading. As a result, they are likely to fall further and further behind their peers especially as they enter middle school where the amount of reading increases dramatically (Stanovich, 1986).

While the instructional intervention for both these students must focus on fluency, their needs are somewhat different. Clearly both need to engage in more reading of relatively easy material to build their automatic word identification and rate of reading, keeping an eye towards maintaining their strength in comprehension. Rate for rate's sake is not the aim. Steven, however, would also benefit from more explicit instruction in decoding multisyllabic words.

Clusters 7 & 8 : Slow Word Callers (17%)

The students in these clusters are word callers, like those in Clusters 1 and 2. However, these students are not automatic; they lack fluency. Overall, these clusters best represent the pattern suggested from the overall descriptive data: accurate readers who are both slow and struggle with meaning. A high percentage of students in this group (50%) scored at Level 1 on the WASL, indicating "little or no demonstration of the prerequisite knowledge and skills that are fundamental for meeting

the standard." (Washington Commission of Student Learning, 1998). For both Joey and José, students in these clusters, this lack of prerequisite skills appears to be in the areas of fluency and meaning. Interestingly, it is not word identification.

Joey's comprehension difficulties are evident from multiple sources: his receptive vocabulary (PPVT-R) was 74, almost 2 standard deviations below average for his age (1st grade equivalency), and his comprehension on the QRI-II averaged 2nd grade. His performance in word identification tells a different story, however. Joey scored at the 9th grade equivalency on word attack and 6th grade equivalency for word identification on the Woodcock Johnson tests, indicating well-developed decoding abilities. He could effectively use word identification skills to read both nonsense words and real words in isolation and in context. In spite of this, Joey was slow.

Joey averaged 70 words per minute when reading 4th grade material from the WASL and the QRI-II. Yet, as we might expect from his strong decoding, he was accurate. That is, Joey didn't stumble on words or have to self-correct; he reached 95% accuracy the first time he saw the words. So, Joey's slow rate cannot be attributed to difficulty with decoding.

José's profile is similar. He is strong in word identification, accurate, and slow. He averaged 56 words per minute on the state test and 63 words per minute when reading from the 4th grade passages on the QRI-II. Like Joey, these rates are considerably below the mean of this group of struggling readers and the NAEP standards (Pinnell, et al., 1995).

José also struggles with comprehension but unlike Joey, José doesn't appear to have difficulty with word meanings. Although Spanish is his first language, his receptive vocabulary is average for a student his age. Instead, comprehension difficulties and his slow pace may be a function of his familiarity with complex reading material in his second language. He may be slowing down to be sure he understands what he reads—a good strategy for him.

Both these boys are slow and struggling with comprehension. In Joey's case, difficulty with word meanings likely contributes, and in José's, second language issues may play a role. Decoding is not an issue for either boy. On one hand, slowing down to attend to comprehension is a good strategy. On the other hand, a slow rate alone can impede comprehension, making it difficult to retain meaning over longer reading selections. A closer look at these students would yield useful information to disentangle their difficulties.

Clusters 9 & 10: Disabled Readers (9%)

Surprisingly, less than 10% of the sample fell into these categories. These students were low in all three areas: word identification, fluency, and meaning. Most telling, however, is their dramatic difficulty with word identification which placed most of these students at a 1st grade level or below in this area. With word identification so low, fluency and comprehension difficulties are likely to follow, resulting in students who struggle on all fronts. Not surprisingly then, the majority of the students in these clusters performed at Level I on the state 4th grade reading assessment, a level indicating a lack of basic skills. Clearly, these students need in-depth intervention at a level beyond what most classroom teachers can provide.

Discussion and Conclusion

Our research probed beneath the surface of students' failing scores on a typical reform-oriented state reading assessment. What we found likely will not surprise reading specialists, but it *is* likely to surprise policymakers, administrators, and some classroom teachers, and it has important implications for their efforts to improve student learning. The evidence here clearly demonstrates that students fail state reading tests for a variety of reasons.

Students who failed a state assessment of reading exhibited several distinctive patterns of performance that contributed to their below-standard scores. Even subtest results based on group averages in meaning, word identification, and fluency, obscured patterns of individual student performance that were markedly different from the group results. In short, reading failure is multifaceted and it is individual. Beneath each failing score is a pattern of performance that holds the key to improved reading instruction and, consequently, improved reading ability.

The results of this study offer direction to states, school districts, and schools for policy that is likely to improve student achievement rather than merely measure it. Briefly, we offer guidance, based on our findings, for policymakers in five areas: instruction; multiple indicators; alignment among standards, assessment and instruction; allocation of resources; and evaluating reform.

Five Potential Policy Levers

Instruction

First, most obviously, we need to question state and district policies that mandate specific instructional strategies or curriculum programs for all failing students. Mandating phonics instruction, for example, for all students who fell below proficiency would have missed large numbers of students (more than 50% in our sample) whose word identification skills were fairly strong yet who struggled with comprehension, fluency, or language. Similarly, an instructional approach that didn't address fluency and building reading stamina of longer, more complex text or that didn't provide sufficient reading material at a range of levels would have missed almost 70% of the students who demonstrated difficulty with fluency.

Instead of mandating particular types of instruction from outside the classroom that are based on average scores of a broad spectrum of students, we suggest that policies should support educators' efforts to probe beneath the surface of test scores in order to determine students' needs and instructional approaches most likely to be effective. The point is to guard against superficial interpretations and responses to test scores. Such policies might take several forms. For example, in addition to requiring action plans—as is common in many state departments of education—districts and schools could be required to engage in, and report on, self-study. That is, instead of simply requiring districts to quickly identify strategies or intervention programs to address "problems" exposed by test scores, states could support schools in question raising and investigation to inform those plans (e.g. Which students did we predict would do well on the test and did not? What contributed to their difficulty? What do we know about the progress of our lowest performing students? What evidence do we have about their learning? What instructional resources and approaches do we have in place for a particular group of students? What do we, as teachers, need to know more about? (Wixson, Valencia, & Lipson, 1994). Such an

approach would require gathering more data and conducting a finer grained analysis of the problem. Ultimately, it would result in more appropriate action plans, better instruction, and improved student learning.

Multiple Indicators

A related suggestion is for policies that require multiple indicators of achievement, a stance that has long been advocated but rarely implemented (Linn, 2000; National Council on Education Standards and Student Testing, 1992). There are varying interpretations of just what multiple indicators might entail—multiple instruments across disciplines, multiple test formats, multiple opportunities for students to demonstrate performance, multiple measures of context and background factors (Cohen & Hill, 2000; Lewis, 2001). At a minimum, our work suggests a need for complex indicators of student performance in the targeted subject area—in this case, multifaceted indicators of students' reading abilities. This is the kind of information that can only come from classroom-based assessments (Shepard, 2000).

Some states, such as Vermont, Kansas, and Washington, have successfully implemented policies targeted at classroom-based assessment strategies in reading at some grade levels. For example, Washington and Vermont require and support all teachers to administer individual reading assessments for their 2nd grade students. The key to these policies is that the required assessments, similar to one used in this study, are: diagnostic; tailored to individual student needs; individually administered, scored, and analyzed by classroom teachers; and supported with professional development to help teachers match effective, research-based instruction to the results of the assessments. This is educative policy at its best (Cohen & Barnes, 1993a). However, simply administering these classroom-based assessments is not what improves teaching and learning. In fact, policies that require teachers to administer a standard battery of classroom assessments and report the scores would miss the point. Instead, mandated assessments that are useful require teacher knowledge and decision-making and, most importantly, are perceived by teachers as helpful in everyday instruction (Au & Valencia, 1997; Black & Wiliam, 1998; Place, in press). Policies directed at classroom-based assessments, such as these, could be useful at all grade levels, not just primary grades. The form of classroom assessments would vary depending on grade level and student abilities, but the point of the policy would be to supplement state test data with classroom-based diagnostic assessments and professional development to match instruction to assessment results. Such policies acknowledge the complexity of learning and teaching and the ever-changing nature of students in our classrooms each year (McLaughlin, 1987).

Alignment Among Standards, Assessment, and Instruction

Our findings also beg for increased understanding of alignment between standards, assessment, and instruction. Alignment is certainly a centerpiece of standards-based reform, and, most of the time, it makes good sense. However, sometimes it can be oversimplified and inadvertently lead to inappropriate instruction. More specifically, aligning instruction with state assessments may help teachers focus on what is tested (in this case, comprehension of different types of texts) but it will not address the skills and strategies that underlie such competence. Alignment of instruction to a particular grade level of standards for all students, without regard to individual skills or strategies will leave children behind. Assuring, for example, that 4th grade teachers are teaching the 4th grade content standards does not assure they are providing appropriate instruction for all students. To be sure, some

students would benefit from instruction and practice reading material that is at a lower grade level, and some would benefit from more advanced curriculum. It is not that grade-level standards or expectations are unimportant or that aligning instruction, assessment, and standards is wrong. However, for many struggling students, grade-level standards are goals rather than immediate needs. The teacher's challenge is to bring the students to the point where those grade level goals are within reach.

We concur with Spillane & Jennings (1997) that it requires more than policy targeted at alignment to get at the more "difficult-to-reach" dimensions of classroom practice. Not only must teachers have the understandings and pedagogical tools to translate these assessments and content standards into good instruction, but they must also reach beneath and beyond these grade level documents. State and district policies can support such practices by providing teachers with curriculum frameworks that clearly articulate both grade level expectations and developmental perspectives on teaching and learning in specific disciplines. The policy and implicit message from states and school districts must be unambiguous: The aim is not simply to teach the standards, it is to teach the students. When 4th grade teachers receive or are told to teach only 4th grade standards or to prepare all students for the 4th grade test, for example, without consideration of student needs, the intent of reform and student achievement are short-circuited.

Allocation of Resources

This study also has implications for local resource policies on curriculum materials, time, instructional organization, and professional development. As teachers provide appropriate instruction for a wide range of students, they will need access to varying levels of reading material, not just for reading instruction, but also for reading in science, social studies, and other content areas. If a student in the 5th grade needs reading instructional material at a 2nd grade level, it is unlikely that the student can read a 5th grade science or social studies text. Textbook adoption policies and procedures in these other subject areas often leave little room for purchasing alternative materials that can be used with students who have a range of reading abilities. Students who are reading substantially below grade level, like those in this study, need appropriate instruction and appropriate materials throughout the school day, not just during reading instruction. They need more instructional time and more precisely targeted instruction than those students who are making adequate progress. For many, this will require additional support beyond the classroom. Yet, with the push on early intervention, many schools have turned their attention to 1st and 2nd grade students, leaving intermediate grade students without supplemental support. The same is true for a great many low-achieving middle and high school students who receive no reading instruction. School and district policies aimed at low-achieving readers need to provide differentiated support for students across abilities and grades.

Perhaps the biggest demand for policy related to resources comes from professional development efforts, the linchpin in improved teaching and learning. Ample evidence suggests that effective professional development is on-going, subject-matter based, and responsive to teachers' beliefs and practices (e.g. Cohen & Ball, 1990; Cohen & Hill; 2000; National Research Council,1999; Firestone, Mayrowetz, & Fairman, 1998; McDonnell, 1994; Spillane & Jennings, 1997). We know too, that professional development must take place in environments that inspire trust and conversation about difficult pedagogical issues (Elmore, Peterson, & McCarthy, 1996; McDonnell, 1994; Wolf, Borko, Elliott,

& McIver, 2000). To this list of effective practices, we would add that professional development is best located in schools and classrooms, where it is more likely to help teachers understand and learn to cope with variability and complexity. This in-situ learning brings static content standards to life and encourages teachers to deal simultaneously with the components of skilled reading, variability in student performance, text difficulty, curriculum materials, instructional strategies, grouping practices and the like. Through classroom-based professional development, teachers learn to be observers and decision-makers. Their knowledge is immediately situated in complex practice.

The challenge, of course, is how to scale-up such an intensive push on professional development. We take our lead from several successful long-term efforts where school district and state policies have provided the framework and ample resources for on-site support of teachers (e.g. Au & Carroll, 1997; Borko, Elliott, & Uchiyama (1999), Elmore & Burney, 1999, Stein, in press). We would suggest a particular emphasis on policies focused toward new and beginning teachers who would benefit from mentoring programs with a strong subject-matter focus and coaches who support their efforts to try instructional strategies that may be different from their more senior colleagues (Grossman, Thompson, & Valencia, in press).

Evaluating Reform

Finally, we caution policymakers, researchers, and administrators to take seriously the complexity behind performance and students' individual variability as they evaluate the success of reform. Armed with this vision, they will need to develop policies that reflect more elaborate and varied images of what counts as progress. State tests will not be sensitive to growth of students who begin very far below grade level or to changes in teaching that may not have immediate impact on students who have long-standing learning difficulties. Even accountability approaches that focus more on school improvement than on current performance (Linn, 2001) may be insensitive indicators of change for the lowest achieving students when broad tests of learning are used. Policies need to support the development of a system of proximal indicators, intermediate indicators, of student learning that can be used to demonstrate student growth, especially for those who cannot yet meet grade level standards. Such indicators are most likely to come from more diagnostic classroom assessments.

In sum, this research is a vivid reminder of the complexity of reading performance and the potential danger of policy that fails to acknowledge this complexity or strategies for dealing with it. No single measure or intervention can possibly meet the needs of all, or even most, of the students who are experiencing reading difficulty. If we take seriously the intent of standards-based reform, which is to improve the quality of instruction and student learning rather than simply to measure and report it, then policies are needed that support people in becoming more knowledgeable, focused, and responsive. We must remember that "below the bar" are individual children with different needs, and behind them are teachers who need policies that support thoughtful teaching and learning.

ENDNOTES

1 If total accuracy were reported simply as a grade level, a student with a score of 95% on the 4th grade narrative passage would receive a score of 4.0, the same score as a student who scored 100% on total accuracy. Conversely, if scores were reported simply as percentages, there would be no way to distinguish a student reading at 2nd grade level with 90% acceptable accuracy from one reading at 5th grade level with 90% acceptable accuracy. To address both these problems, we created a common scale by multiplying percent correct on each measure by the student's instructional level.

2 Studies of rate and fluency have used a wide range of reading selections and methods for determining both rate and expression. Depending on the passages read, average rates for 4th grade students range from 120-170 word per minute (Harris & Sipay, 1990; McCracken, 1970, Taylor, 1965); for 5th grade they range from 120-177 (Harris & Sipay, 1990; McCracken, 1970, Taylor, 1965).

3 Cluster analysis uses either a distance measure or a correlation measure to define similarity within a cluster, depending on the population of interest. A correlation measure of similarity was selected here because it forms clusters based on students' patterns of relative performance across measures. When correlation is used as the similarity measure, then an average linkage algorithm is preferred to create clusters of students whose mean scores are more similar to one another than to any other group. The size of a cluster can range from one subject to a single cluster that includes all the subjects in a study. The number of clusters that best describe a data set is determined by examining the differential statistic represented by a jump in coefficients. This statistic indicates when clusters containing quite dissimilar members are being combined.

REFERENCES

Abelman, C. & Elmore, R.F. (1999). *When accountability knocks, will anyone answer?* CPRE Research Report No. RR-042. Philadelphia: University of Pennsylvania, Consortium for Policy Research in Education.

Adams, M. J. (1990). *Beginning to read: Thinking and learning about print.* Cambridge, MA: MIT Press.

Adams, M.J., Foorman, B.R., Lundberg, I. & Beeler, T (1998). *Phonemic awareness in young children: A classroom curriculum.* Baltimore: Paul H. Brookes Publishing Co.

Aldenderfer, M. & Blashfield, R. (1984). *Cluster analysis.* Beverly Hills: Sage.

Allington, R.L. (2001). *What really matters for struggling readers.* New York: Longman.

Allington, R. L (2000). *Effects of reading policy on classroom instruction and student achievement.* Report Series 13011. Albany, NY: National Research Center on English Learning & Achievement.

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

Allington, R.L. & Walmsley, S. (1995).) *No quick fix : rethinking literacy programs in America's elementary schools*. New York: Teachers College Press.

Amrein, A.L. & Berliner, D.C. (2002). High-stakes testing, uncertainty, and student learning. *Education Policy Analysis Archives, 10*(18). Available: http://epaa.asu.edu/epaa/v10n18/.

Anderson R.C. & Freebody, P. (1981). Vocabulary knowledge. In J.T. Guthrie (Ed.), *Comprehension and teaching: Research perspectives* (pp. 71-117). Newark, DE: International Reading Association.

Au, K. H. and J. H. Carroll (1997). Improving literacy achievement through a constructivist approach: The KEEP Demonstration Classroom Project. *Elementary School Journal, 97*(3): 203-221.

Au, K.H. & Valencia, S.W. (1997). The complexities of portfolio assessment. In D. Hansen & N. Burbules (Eds.), *Teaching and its predicaments* (pp. 123-144). Boulder, CO: Westview.

Black, P. & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education, 5*(1), 7-74.

Borko, H., Elliott, R., & Uchiyama, K. (1999). *Professional development: A key to Kentucky's reform effort*. CSE Tech Rept 512. CRESST: University of Colorado at Boulder.

Brown. R.G. (1991) *Schools of thought : How the politics of literacy shape thinking in the classroom*. San Francisco : Jossey-Bass.

Bus, A.G., & Van Izendoorn, M.H. (1999). Phonological awareness and early reading: A meta-analysis of experimental training studies. *Journal of Educational Psychology, 91*, 403-414.

Callahan, S. (1997). Tests Worth Taking?: Using Portfolios for Accountability in Kentucky. *Research in the Teaching of English, 31*, 295-336.

Carver, R.P. (1990). *Reading rate: A review of research and theory*. San Diego: Academic.

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

Clay, M. M. (1993a). *An observation survey of early literacy achievement*. Portsmouth, NH: Heinemann.

Clay, M. M. (1993b). *Reading recovery: A guidebook for teachers in training.* Portsmouth, NH: Heinemann.

Cohen, D.K. & Ball, D.L. (1990). Relations between policy and practice: A commentary. *Educational Evaluation and Policy Analysis, 12*, 249-256.

Cohen, D. K. & Barnes, C. A. (1993a) Pedagogy and policy. In Cohen, D.K., McLaughlin, M.W., & Talbert, J.E. (Eds.). *Teaching for understanding: Challenges for policy and practice* (pp. 204-239). San Francisco: Jossey-Bass.

Cohen, D. K. & Barnes, C. A. (1993b) Pedagogy and policy. In Cohen, D.K., McLaughlin, M.W., & Talbert, J.E. (Eds.). *Teaching for understanding: Challenges for policy and practice* (pp. 240-276). San Francisco: Jossey-Bass.

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

Cohen, D. K., & Spillane, J. (1992). Policy and practice: The relations between governance and instruction. In G. Grant (Ed.), *The Review of Research in Education* (pp. 3-49). Washington, DC: American Educational Research Association.

Collier, V. P. (1987). Age and rate of acquisition of second language for academic purposes. *TESOL Quarterly, 21*, 617-641.

Collier, V. P. (1989). How long? A synthesis of research on academic achievement in a second language. *TESOL Quarterly, 23*, 509-531.

Cummins, J. (1991). The development of bilingual proficiency from home to school: A longitudinal study of Portuguese-speaking children. *Journal of Education, 173*, 85-98.

Davis, F.B. (1968). Research in comprehension in reading. *Reading Research Quarterly, 3*, 499-545.

Dowhower, S.L. (1987). Effects of repeated reading on second-grade transitional reader' fluency and comprehension. *Reading Research Quarterly, 22*, 389-406.

Duffy, G. G. & Hoffman, J. V. (1999). In pursuit of an illusion: The flawed search for a perfect method. *The Reading Teacher, 53*, 10-15.

Dunn, L.M. & Dunn, L.M. (1981). *Peabody Picture Vocabulary Test-Revised.* Circle Pines, MN: American Guidance Service.

Edmondson, J. & Shannon, P. (2002). The will of the people. *The Reading Teacher, 55*, 452-454.

Education Week (2001). Quality Counts 2001. *Education Week, 20*(17).

Ehri, L.C. (1992). Reconceptualizing the development of sight word reading and its relationship to recoding. In P.Gough, L.Ehri, & R. Treiman (Eds.), *Reading Acquisition* (pp. 107-143). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Elmore, R.F. & Burney, D. (1998). *The challenge of school variability: Improving instruction in New York City's Community School District #2*. Philadelphia: Center for Policy Research in Education, University of Pennsylvania.

Elmore, R.F., Peterson, P. & McCarthy, S.J. (1996). *Restructuring in the classroom: Teaching, learning, and school organization*. San Francisco: Jossey-Bass Publishers.

Everitt, B. (1974). *Cluster analysis.* New York: Wiley.

Firestone,W.A., Mayrowetz, D., & Fairman, J. (1998). Performance-based assessment and instructional change: The effects of testing in Maine and Maryland. *Educational Evaluation and Policy Analysis, 20*, 95-114.

Fox, K.R. (2001). *The formation of the Governor's reading initiative: A descriptive policy study in reading*. Paper presented at the annual meeting of the National Reading Conference, San Antonio.

Goldberg, G.L. & Kapinus, B. (1993). Problematic responses to reading performance assessment tasks: Sources and implications. *Applied Measurement in Education, 6*, 281-305.

Goodnough, A. (2001, May 23). Teaching by the book, no asides allowed. *The New York Times.* (Available: http://www.nytimes.com).

Gough, P. B., Hoover, W., & Peterson, C.L. (1996). Some observations on the simple view of reading. In c. Cornoldi & J. Oakhill (Eds.), *Reading comprehension difficulties* (pp. 1-13). Mahwah, NJ: Lawrence Earlbaum Associates.

Grossman, P.L., Thompson, C., & Valencia, S.W. (in press). Focusing the concerns of new teachers: The district as teacher educator. In A. Hightower, M.S. Knapp, J.A, Marsh, & M.W. McLaughlin (Eds.). *School districts and instructional renewal*. New York: Teachers College Press.

Harris, R. (2000, May 29). *Genetic research.* Morning Edition, National Public Radio.

Harris, A.J. & Sipay, E.R. (1991). *How to increase reading ability (9th ed.).* New York: Longman.

Hiebert, E.H. & Taylor, B. M. (Eds.)(1994). *Getting reading right from the start: Effective early literacy interventions*. Boston: Allyn & Bacon.

Hochstatter, H. (1998, February 25). Johnny can't read because Johnny needs phonics. *The Seattle Times.*

Hoien, T., Lundberg, I., Stanovich, K.E., Bjaalid, I-K (1995). Components of phonological awareness. *Reading and Writing, 7*, 171-188.

Jenkins, J. R., Johnson, E., & Hileman, J. (2001). *When is reading also writing: Sources of individual differences on the new reading performance assessments*. Unpublished manuscript. University of Washington.

Klein, S.P., Hamilton, L.S., McCaffrey, D.F., & Stecher, B.M. (2000). What do test scores in Texas tell us? *Education Policy Analysis Archives, 8*(49). Available: http://epaa.asu.edu/epaa/v8n49/

Ko, M. (1997, March 8). Locke, GOP reach deal to pass phonics-based education bill. *The Seattle Times.*

LaBerge, D. & Samuels, S. J. (1974). Toward a theory of automatic information processing in reading. *Cognitive Psychology, 6*, 293-323.

Leslie, L. & Caldwell, J. (1995). *Qualitative Reading Inventory-II.* New York: HarperCollins College Publishers.

Lewis, A. (2001). *Benchmarks for accountability: Are we there yet?* CSE Technical Report 547. National Center for Research on Evaluation, Standards, and Student Testing. Los Angeles: University of California.

Linn, R.L. (2001). *Reporting school quality in standards-based accountability systems*. CRESST Policy Brief 3. Los Angles: National Center for Research on Evaluation, Standards, and Student Testing.

Linn, R. L. (2000). Assessment and accountability. *Educational Researcher, 29*(2), 4-16.

Lipson, M. Y. & Wixson, K. K. (1997). *Assessment and instruction of reading and writing disability: An interactive approach*, 2nd edition. New York: Longman.

Marsh, J. (2000). *Connecting districts to the policy dialogue: A review of literature on the relationship of districts with states, schools, and communities*. A working paper. Seattle: Center for the Study of Teaching and Policy.

McCracken, R. (1970). *The teaching of reading: A primer*. Klamath Falls, OR: Klamath.

McDaniel, J.E. & Miskel, C.G. (2001). *The nature of influence in national reading policymaking: The role of beliefs and the importance of collaboration*. Paper presented at the annual meeting of the National Reading Conference, San Antonio.

McDaniel, J.E., Sims, C.H., & Miskel, C.G. (2001). The national reading policy arena: Policy actors and perceived influence. *Education Policy 15*(1), 92-114.

McDonnell, L. (1994). Assessment policy as persuasion and regulation. *American Journal of Education, 102*, 394-420.

McDonnell, L. & Choisser, C. (1997). *Testing and teaching: Local implementation of new state assessments*. Los Angeles: National Center for Research on Evaluation, Standards, and Student Testing, University of California.

McLaughlin, M. W. (1987). Learning from experience: Lessons from policy implementation. *Educational Evaluation and Policy Analysis, 9*, 171-178.

McNeil, L.M. (2000). *Contradictions of school reform: Educational costs of standardized testing*. New York: Routledge.

Nagy, W. E. (1988). *Teaching vocabulary to improve reading comprehension*. Urbana, IL: ERIC Clearinghouse on Reading and Communication Skills and the National Council of Teachers of English.

National Center for Education Statistics. (1996). *NAEP 1994 reading report card for the nation and the states.* Washington: Office of Educational Research and Improvement, U.S. Department of Education.

National Council on Education Standards and Testing. (1992). *Raising standards for American education*. Washington, DC: U.S. Government Printing Office.

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

National Research Council (1999). *Testing, teaching, and learning: A guide for states and school districts.* Committee on Title I testing and assessment, R.F. Elmore & R. Rothman, Ed. Board on Testing and Assessment, Commission on Behavioral and Social Sciences and Education. Washington D.C.: National Academy Press.

Nolen, S. B., Haladyna, T. M., & Haas, N. S. (1992). Uses and abuses of achievement test scores. *Educational Measurement: Issues and Practice, 11*(2), 9-15.

Paterson, F.R.A. (2000). The politics of phonics. *Journal of Curriculum and Supervision, 15*, 179-211.

Perfetti, C. A. (1985). *Reading ability*. New York: Oxford University Press.

Pinnell, G. S., Pikulski, J. J., Wixson, K. K., Campbell, J. R., Gough, P. B. & Beatty, A. S. (1995). *Listening to children read aloud*. Washington, DC: US Department of Education.

Place, N.A. (in press). Policy in action: The influence of mandated early reading assessment on teachers' thinking and practice. *Fiftieth Yearbook of the National Reading Conference.*

Popham, J.W., Cruse, K.L., Rankin, S.C,, Sandifer, P.D., & Williams, P.L. (1985). Measurement-driven instruction: It's on the road. *Phi Delta Kappan, 66*, 628-634.

Resnick L.B. & Resnick, D.P. (1992). Assessing the thinking curriculum: New tools for educational reform. In B.R. Gifford & M.C. O'Connor (Eds.), *Changing assessments: Alternative views of aptitude, achievement, and instruction*. Boston: Kluwer Academic Publishers.

Romesburg, H.C. (1984). *Cluster analysis for researchers*. Belmont, CA: Lifetime Learning Publications.

Scarborough, H.S., Ehri, L.C., Olson, R.K., Fowler, A. (1998). The fate of phonemic awareness beyond the elementary school years. *Scientific Studies of Reading, 2*, 115-142.

Shepard, L. A. (1991). Will national tests improve student learning? *Phi Delta Kappan, 72*, 232-238.

Shepard, L.A. (2000). The role of assessment in a learning culture. *Educational Researcher, 29*, 4-14.

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

Spear-Swerling, L., & Sternberg, R.J. (1994). *Educational implications of an interactive model of reading disability*. Paper presented at the Annual Meeting of the American Educational Research Association (San Francisco, CA, April 18-25, 1995).

Speece, D.L. (1994). Cluster analysis in perspective. *Exceptionality, 5*, 31-44.

Spillane, J. P. (2000). Cognition and policy implementation: District policymakers and the reform of mathematics education. *Cognition and Instruction, 18*(2), 141-179.

Spillane, J. P. (1996) School districts matter: Local educational authorities and state instructional policy. *Educational Policy, 10*, 63-87.

Spillane, J.P. & Jennings, N.E. (1997). Aligned insturctioanl policy and ambitious pedgagoy: Exploring instructional reform from the classroom perspective. *Teachers College Record, 98*, 449-481.

Spillane, J.P. & Thompson, C.L. (1997). Reconstructing conceptions of local capacity: The local education agency's capacity for ambitious instruction reform. *Education Evaluation and Policy Analysis, 19*, 185-203.

Stahl, S.A. & Murray, B. (1998). Issues involved in defining phonological awareness and its relation to early reading. In J.L. Metsala & L.C. Ehri (Eds.), *Word recognition in beginning literacy*. (pp. 65-88). Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.

Standerford, N. S. (1997). Reforming reading instruction on multiple levels: Interrelations and disconnections across the state, district, and classroom levels. *Educational Policy, 11*, 58-91.

Stanovich, K.E. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly, 21*, 360-407.

Stanovich, K.E. (1994). Romance and reality. *The Reading Teacher, 47*, 280-290.

Stein, M.K., & D'Amico, L. (in press). District as a Professional Learning Laboratory. In A. Hightower, M.S. Knapp, J.A, Marsh, & M.W. McLaughlin (Eds.). *The school district and instructional renewal: Opening the conversation*. New York: Teachers College Press.

Taylor, C. S. (1998). *Assessment of student learning: Grade 4 technical report*. Olympia, WA: Office of the Superintendent of Public Instruction.

Taylor, S.E. (1965). Eye movements in reading: Facts and fallacies. *American Educational Research Journal, 34*, 193-205.

Valencia, S.W. & Wixson, K.K. (2001). *Literacy policy and policy research that makes a difference*. In J.V. Hoffman, D. L. Schallert, C.M. Fairbanks, J. Worthy, B. Malock (Eds.). *The Fiftieth Yearbook of the National Reading Conference* (pp. 21-43). Chicago: National Reading Conference.

Wagner, R.K., Torgesen, J.K., & Rashette, C.A. (1999). *The Comprehensive Test of Phonological Processing*. Austin, TX: Pro Ed.

Wagner, R.K., Torgesen, J.T. & Rashette, C.A. (1994). The development of reading-related phonological processing abilities: New evidence of bi-directional causality from a latent variable longitudinal study. *Developmental Psychology, 30*, 73-78.

Washington Commission on Student Learning (1998). *Understanding the new 4th and 7th grade test: a guide for parents*. Olympia: Office of the superintendent of Public Instruction.

Washington State Legislature (1998). *Successful Readers Act, E2SSB 6509*, Winter interim session, Olympia, Washington.

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

Wolf, S.A., Borko, H., Elliott, R.L., & McIver, Monette (2000). “That dog won’t hunt!”: Exemplary school change efforts within the Kentucky reform. *American Educational Research Journal, 37,* 349-393.

Woodcock, R.W. & Johnson, M.B. (1989). *Woodcock-Johnson Psychoeducational Battery-Revised.* Allen, TX: Psychological Corporation.

CTP Occasional Papers

The Center's Occasional Paper Series addresses topics that are timely and intimately connected to the Center's agenda. Papers in the series—some by Center members, others by researchers elsewhere —include conceptual work, research syntheses, discussions of policy issues, and reports of empirical studies undertaken under other auspices. Occasional Papers are externally reviewed and revised prior to publication by the Center. Along with CTP Research Reports, Working Papers, and Policy Briefs, these papers are available for download from the Center's website: **www.ctpweb.org**

Center Affiliates

American Association of Colleges of Teacher Education
American Federation of Teachers
Council for Chief State School Officers
National Alliance of Business
National Association of Secondary School Principals
National Board for Professional Teaching Standards
National Council for Accreditation of Teacher Education
National Council of Teachers of English
National Education Association
National School Boards Association
National Staff Development Council
National Urban League
American Association of School Administrators
Association for Supervision and Curriculum Development
International Reading Association
National Association of Elementary School Principals
National Association of State Boards of Education
National Conference of State Legislatures
National Council for the Social Studies
National Council of Teachers of Mathematics
National Governors' Association
National Science Teachers Association
National Urban Coalition
Teachers Union Reform Network

Center Team

Principal Investigators and Co-Principal Investigators

UNIVERSITY OF WASHINGTON
Michael Knapp, Center Director
James Banks
Elizabeth Dutro
Margaret Plecki
Sam Wineburg
Sheila Valencia

STANFORD UNIVERSITY
Linda Darling-Hammond
Pamela Grossman
Milbrey McLaughlin
Joan Talbert

UNIVERSITY OF MICHIGAN
Deborah Loewenberg Ball
David Cohen
Edward Silver

UNIVERSITY OF PENNSYLVANIA
Thomas Corcoran
Richard Ingersoll

Researchers at Other Institutions
Barnett Berry, University of North Carolina
Robert Floden, Michigan State University
David Monk, Pennsylvania State University
Jon Snyder, Bank Street College
Judy Swanson, Education Matters, Inc.
Suzanne Wilson, Michigan State University

Contact Information

Michael S. Knapp, Center Director
Miller Hall M201, College of Education
University of Washington, Box 353600
Seattle, WA 98195-3600
email: mknapp@u.washington.edu

Michele C. Ferguson, Center Manager
Miller Hall 203C, College of Education
University of Washington, Box 353600
Seattle, WA 98195-3600
Phone: (206) 221-4114
FAX: (206) 616-6762
email: ctpmail@u.washington.edu

Sally Brown, Communications Director
Miller Hall 404B, College of Education
University of Washington, Box 353600
Seattle, WA 98195-3600
Phone: (206) 543-5319
FAX: (206) 616-6762
email: salbrown@u.washington.edu

Web Address
http://www.ctpweb.org