How Long Is Long Enough? Fourth Grade

English Language Learners' Scores on a State's Test and Lengths of Stay

Claudia Olmstead

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

Doctor of Philosophy

University of Washington

2004

Program Authorized to Offer Degree:
College of Education

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.

ProQuest LLC
789 E. Eisenhower Parkway
PO Box 1346
Ann Arbor, MI 48106-1346

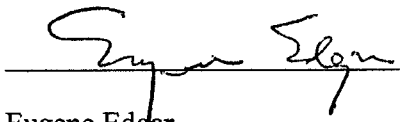
University of Washington
Graduate School

This is to certify that I have examined this copy of a doctoral dissertation by

Claudia Olmstead

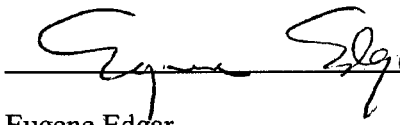
And have found that it is complete and satisfactory in all respects,
and that any and all revisions required by the final
examining committee have been made.

Chair of Supervisory Committee:

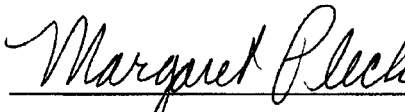


Eugene Edgar

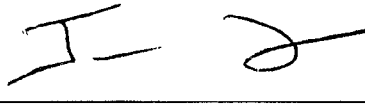
Reading Committee:



Eugene Edgar



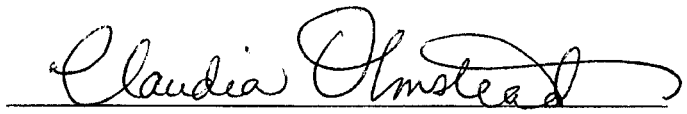
Margaret Plecki



Tom Stritikus

Date: 8-16-04

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, 3000 North Zeeb Road, Ann Arbor, MI 48106-1346, to whom the author has granted "the right to reproduce and sell (a) copies of the manuscript in microform and/or (b) printed copies of the manuscript made from microform."

Signature 

Date 8-16-04

University of Washington

Abstract

How Long Is Long Enough? Fourth Grade
English Language Learners' Scores on a State's Test and Lengths of Stay

Claudia Olmstead

Chair of the Supervisory Committee:
Dr. Eugene Edgar
College of Education

The point at which English Language Learners can participate meaningfully in new mandated state assessments remains a topic of debate. Researchers identify several areas of concern. A wealth of studies from the last 25 years indicate ELLs need from 4 to 10 years to achieve in English at the academic level of native English-speaking peers (Collier, 1992; Collier & Thomas, 2004; Cummins, 1981; De Avila, 1997; Hakuta, Butler & Witt, 2000; Ramirez, Yuen & Ramey, 1991; Thomas & Collier, 1997, 2002). The validity of testing students' content knowledge in English while they are still acquiring English has been questioned (August & Hakuta, 1997; Hakuta, 2001; Holmes & Duron, 2000; LaCelle-Peterson & Rivera, 1994; Olson & Goldstein, 1997; Rivera & Stansfield, 1998). Also, the appropriateness of an assessment that was developed for native English speakers is argued (Abedi & Dietel; Coucil of Chief State School Officers, 2000; LaCelle-Peterson & Rivera, 1994).

In this descriptive study, testing data for fourth grade ELLs from five neighboring Northwest school districts were gathered over 3 testing years. The data were disaggregated by number of years in district. It was possible to compare test results for students in their districts for 1 year, 2 years, 3 years, 4 years, and 5 years. Also, test data were dissagregated for students formerly in a district ELL program but redesignated as fluent English proficient. Results showed slow, uneven progress for ELLs at consecutive years in district. In Listening and Reading, more progress occurred than in Mathematics and Writing. Yet most ELLs still could not meet standard after 5 years. However, redesignated students had more success; they achieved at levels closer to English only peers.

A further qualitative component examined administrators' reactions to the study data. Superintendents, elementary education directors and bilingual program directors from the five districts were interviewed after receiving a report summarizing the findings. After reviewing the report, all administrators but one had alternative suggestions for more equitable inclusion of ELLs in the state assessment system.

TABLE OF CONTENTS

Page

LIST OF FIGURES

LIST OF TABLES

Table number Page

ACKNOWLEDGEMENTS

I would like to express sincere appreciation to the Department of Education for their support in helping me complete my studies. Living across the state presented some extra challenges, and I am grateful for all the help I received. Especially, I thank Professor Eugene Edgar for all of his assistance and encouragement. I thank the staff in the Office of Admissions and Academic Support, who helped me more times and in more ways than I can count. Also, I acknowledge with gratitude the strong network of support and care provided by my family and friends.

DEDICATION

To my mother, Barbara Neely.

To my children, Sam and Sara Hoober.

Chapter 1

Introduction

Background

National education reform requires annual statewide testing of all U.S. students in grades three through eight, completely operational by the 2005-2006 school year (*No Child Left Behind Act of 2001*, 2002). The tests are known as ***high stakes assessments;*** they affect decisions about teacher competency and school funding as well as whether students go on to the next grade or graduate from high school. In fact, half the nation's states will use test scores as criteria for high school graduation (Voke, 2002).

Educators and policymakers question the most appropriate way to include the steadily growing group of English language learners (Council of Chief State School Officers, 2000; LaCelle-Peterson & Rivera, 1994; National Research Council, 2000). These students speak native languages other than English and are in the process of acquiring English proficiency. Also referred to as limited English proficient (LEP), this group doubled in size from 1990 to 2000 (Allen, 2002) and now account for at least 9.6% of the U.S. student population (Kindler, 2002). By the 2030's, it is estimated that this group will grow to 40% of U.S. students (Thomas & Collier, 2002). In 2001-2002, 43% of U.S. school districts enrolled English language learners (ELLs), and 43% of U.S. teachers reported teaching ELLs (U.S. Department of Education, 2004).

Most educators welcome the inclusion of English language learners in educational reform. Past educational reforms often overlooked ELLs (Calderon, 1997; Council of

Chief State School Officers, 2000; McDonnell & Hill, 1993; McLaughlin & McLeod, 1996; Miramontes, Nadeau & Commins, 1997). Abilities of ELLs have regularly been underrated; they have been under-challenged and disproportionately assigned to lower curricular tracks (LaCelle-Peterson & Rivera, 1994; McLaughlin & McLeod, 1996; Miramontes et al., 1997). The new focus on achievement for all groups accompanied by data analysis can help identify the particular programs and instructional methods most effective for ELLs (August, Hakuta & Pompa, 1994; Collier & Thomas, 2004).

However, educators and researchers identify several major issues concerning statewide assessments and English language learners. They include: the problems of classifying all ELLs, a very inconstant group, as one subgroup as mandated by the new *No Child Left Behind Act* (henceforth NCLBA) guidelines; the problems of testing in English when learners are not English-fluent; the length of time required to acquire English academic fluency; and the suitability of state assessments to measure ELL achievement.

Problems of classifying ELLs as one subgroup. Because schools and districts are required to report ELLs as one subgroup, only large generalizations can be made. Continual new arrivals typically score lower than students who have been in their district for several years, while students redesignated as English-proficient who typically score the highest are usually removed from the ELL subgroup and placed with the native English speakers (Abedi & Dietel, 2004). Further, other persistent variables accounting for wide differences within this group are masked when ELLs are grouped as one mass. These include educational background, amount of literacy in the first language, parent educational level, differing state and district criteria for ELL identification, differences in

state assessments, family socioeconomic level and quality of educational program (Abedi, 2004; Abedi & Dietel, 2004; Council of Chief State School Officers, 2000; Thomas & Collier, 1997). Also, in the reporting of meeting or not meeting standard, small ELL groups from smaller districts are not weighted differently from large ELL populations from larger districts. Thus, a smaller group can be skewed by the performance of a few students while a larger group will reflect a stronger group average (Abedi, 2004; Abedi & Dietel, 2004; Linn, 2003; Linn, R.L., Baker, E.L. & Herman, J.L., 2002). As Abedi and Dietel (2004) note, ELLs are a very unstable subgroup. Therefore, subgroup assessment information has limited usefulness: it does not reflect the range of ELL performance. Nor does it help schools, districts or states to examine their own effectiveness.

However, Secretary of Education Rod Paige (U.S. Department of Education, 2004) recently announced two new provisions to NCLBA that may add some stability to the subgroup's achievement data. First, states do not have to include ELL testing data for ELLs in their first year of residence. Second, states can now continue including achievement data from students exiting their state's language assistance program for two subsequent years in the ELL subgroup. These changes became effective immediately. It is too early yet to gauge their effect on test scores.

Problems of testing in English. Standardized testing of ELLs in English measures English proficiency as well as skills or content knowledge, compromising validity and accuracy (August & Hakuta, 1997; Hakuta, 2001; Holmes & Duron, 2000; LaCelle-Peterson & Rivera, 1994; Olson & Goldstein, 1997; Rivera & Stansfield, 1998). The ELL group continually scores 20 to 30% behind native English-speaking students (Abedi &

Dietel, 2004). In some studies, English language learners demonstrate higher achievement when tested in their primary language (Escamilla, Mahon, Riley-Bernal & Rutledge, 2003) and when language complexity of test items is reduced (Abedi & Lord, 2001; Abedi, Lord & Plummer, 1997).

Length of time to gain English academic fluency. Numerous studies from the last 25 years indicate it takes ELLs 4 to 10 years to score at the level of English fluency of their native English-speaking peers (Collier, 1992; Collier & Thomas, 2004; Cummins, 1981; De Avila, 1997; Hakuta, Butler & Witt, 2000; Ramirez, Yuen & Ramey, 1991; Thomas & Collier, 1997, 2002). While students can become fluent in conversational English within a year or two, academic English fluency takes them several years more (Collier, 1992; Cummins, 1981; De Avila, 1997; Hakuta et al., 2000). A more recent study, examining more than 2 million student records over 18 years, finds that 6 to 8 years are consistently necessary to reach native-English speaking performance levels (Collier & Thomas, 2004).

Suitability of state assessments to measure ELL achievement. A 2000 position statement from the American Evaluation Association condemns use of a single evaluative measure for high stakes decisions for any students. They warn that for ELLs, tests in English may not reflect their capabilities; further, processes of test development, review and validation should involve these students as well as all particular groups of learners. Yet ELLs have largely been excluded from the test development process (August et al., 1994; Council of Chief State School Officers, 2000). Inclusion of ELLs in the test design and piloting process is necessary, the Council of Chief State School Officers (henceforth CCSSO, 2000) argues. Further, tests should be subject to rigorous and expanded bias

reviews that identify items with language difficulty (Abedi and Dietel, 2004; CCSSO, 2000).

Purpose of Study

In Washington State, as in other places, English language learners are a growing group. In the school year 1991-1992, a total of 34,338 ELLs were served in the state's transitional bilingual program; in 2001-2002, it more than doubled to 72,215 students (Office of Superintendent of Public Instruction, 2003a).

The percentage of fourth grade ELLs passing the Washington Assessment of Student Learning for 2001-2002 shows an even larger gap than the typical 20 to 30% gap between ELLs and native English-speaking students on achievement tests (Abedi and Dietel, 2004). On Washington's reading assessment, while 65.6% of all fourth graders met standard, only 24.8% of ELLs met standard. In math, 51.8% of all fourth grade students met standard; 18.2% of the English language learners met standard (Office of Superintendent of Public Instruction, 2002a).

To better understand these data, more detailed analysis of the assessment data is needed. The entire range of ELL test performance needs to be examined (CCSSO, 2000). Hakuta et al. (2000) report findings that "graphically display English proficiency development as a function of time" (p.1). Therefore, students' lengths of stay in district ELL programs are one contributing source of further information.

This study will further disaggregate testing data for fourth grade English language learners by number of years students have spent in their districts- at 1 year, at 2 years, at 3 years, 4 years and 5 years. It will seek a relationship between time in district and test

achievement. To examine the complete ELL range, it will also examine scores of those who have met state criteria for redesignation as fluent English speakers.

Data from five neighboring school districts in southeastern Washington will be examined. While these districts are geographically close, they vary considerably in size, economic base, demographics, resources and educational programs. These districts represent to a degree the variety identified in different parts of the state of Washington by the Office of Superintendent of Public Instruction (henceforth OSPI, 2000a). These five districts will serve as a case study about fourth grade ELL WASL performance in one area of Washington over a 3-year period.

In the following chapter, key literature related to the study focus will be reviewed. Specifically, Chapter 2 will examine research surrounding these major issues: a) the problems of classifying ELLs as one subgroup, b) testing in a non-native language; c) the length of time needed for development of academic English fluency; and d) the suitability of statewide assessments, in particular Washington's, as measures of ELL achievement. Next, Chapter 3 will describe the research methods: the research questions, the setting and sample, and the strategies for data collection and data analysis. Chapter 4 will then present the study's findings. In conclusion, Chapter 5 will provide a discussion of the study's findings and their import.

This study may contribute to a more layered understanding of ELL performance on Washington's Assessment of Student Learning. On a small scale, it will view the alignment of the state's transitional bilingual program's exit criteria with readiness to meet state standards for students in these five districts. This information is important for administrators and policymakers who report on test outcomes for all students and seek

further data analysis and insight. Additionally, it may inform leaders who influence policy. Lastly, two new recent provisions to NCBLA have been announced: a) non-inclusion of testing data for ELLs who have arrived within the last year, and b) inclusion of testing data for students redesignated as fluent English proficient and exited from the bilingual program within the last 2 years. This study may help forecast possible outcomes of these two new provisions on achievement data for these five districts.

Chapter 2

Review of Literature

This study searched for a relationship between length of time in districts for English language learners and their performance on a state achievement test. Additionally, it examined achievement of students previously classified as limited English proficient but redesignated as fluent English proficient (FEP). In this chapter, relevant literature is examined.

This chapter begins with a summary of how the *No Child Left Behind Act* (2002) provides for participation of ELLs. Following that, the chapter discusses the problems of classifying ELLs as one subgroup. The chapter then addresses the problems that arise from high stakes testing of English learning students in a non-native language. The next part of the chapter specifically reviews the literature in the area of the study's focus, the length of time needed to develop English academic fluency skills. The final part of the chapter addresses suitability of current statewide assessments for ELLs and, in particular, the appropriateness of the Washington Assessment of Student Learning (WASL).

No Child Left Behind and ELLs.

In the *No Child Left Behind Act of 2001* (2002, Title III), President George W. Bush specifically addressed ELL achievement:

> (R)esearch has shown that English language learners, when compared with their English-fluent peers, tend to receive lower grades and often score below the average on standardized math and reading assessments. This is partly because federal funding for bilingual education currently has no performance measures

attached to it. Our proposal will give districts more flexibility . . . for effectively transitioning LEP students into English fluency and improving their achievement.

The *No Child Left Behind Act of 2001* (2002) requires annual reports from states, reporting their progress in closing the achievement gap between minorities and disadvantaged students (such as ELLs) and their peers. Under NCLBA, each state develops its own assessment system, with a set of standards for all students, measured by state tests. The states develop their own Adequate Yearly Progress (AYP) requirements, approved by a Peer Review process established by NCLBA. Based on improved student test scores, states must meet these scores to demonstrate success in meeting their own targets.

Under the initial NCLBA guidelines (2002), testing data for ELLs in reading and language arts had to be included after they had attended school in the United States for 3 consecutive years. At first, 44 states permitted from 1 to 3 year exemptions for ELLs from tests, usually based on English proficiency level or time in U.S. (Rivera, Vincent, Haffner, and LaCelle-Peterson, 1997). However, subsequent NCLBA guidelines disallowed these exemptions (U.S. Department of Education, 2002). The later requirements set by NCLBA also included the following: a) Schools and local educational agencies were to be held accountable for all subgroups. Subgroups were defined as the following: each major ethnic and racial group, economically disadvantaged students, students with disabilities, limited English proficient students, and migrant students. At least 95% of each subgroup had to be included in the testing data pool. b) Students in attendance less than a full academic year did not have to be included. c) All students needed to be tested in reading or language arts and mathematics; the state's

definitions of AYP had to be toward a goal of 100% proficiency in both subject areas for all students, in all subgroups, by 2013-14. d) Besides test data, AYP criteria would also include high school graduation rates and could include such other criteria as attendance rates. e) Any Title I school with a subgroup failing to make AYP for two years in a row would be identified as "in need of improvement", and the school would be placed on a timeline for improvement with clearly defined measures (corrective actions) to be taken (see Table 1). f) A state's accountability system had to have validity and reliability.

Table 1.
Corrective actions for failure to meet Adequate Yearly Progress (AYP).

Year	Status	Corrective Action
One	School Improvement, Year One (begins when students fail to make AYP 2 years in a row)	Local educational agency (LEA) must provide technical assistance as the school addresses its academic achievement problem and develops a two-year school improvement plan based on its achievement data. In addition, all students must be offered public school choice.
Two	School Improvement, Year Two	Supplemental educational services must be made available to students of low-income families. LEA continues to assist in implementation of new plan and offers school choice.

Table 1 (continued)

Three	Corrective Action	LEA must take at least one of the following actions: a) replace staff tied to failure to make AYP b) implement new research-based curriculum accompanied by professional development c) decrease school-level management authority d) extend school day or school year e) appoint outside expert as consultant f) reorganize school internally
Four	Restructuring	The LEA must prepare a plan and make arrangements to carry out one of the following: a) reopen school as charter school b) replace principal and staff c) contract with private company to manage school d) allow the state to take over school e) other form of restructuring
Five	Implementation of Restructuring	One of the above arrangements must be implemented by the first day of the school year.

Note: from U.S. Department of Education, 2002 [On-line]. *Paige outlines adequate yearly progress provisions under No Child Left Behind.* Available from http://www.ed.gov/News/ pressreleases/2002/07/07242002.html

In an analysis of the later 2002 NCLBA regulations, Linn (2003) argued that adding subgroups to school, district and state requirements "reduces the possibility that a school will meet its goal, even if all other factors remain equal" (p.2). Linn demonstrated that smaller subgroup sizes are less statistically sound than larger groups and that focusing on subgroups weakens statistical chances of meeting goals. An additional difficulty is that lower achieving schools have greater ground to cover to make the required progress and thus face higher obstacles in comparison to higher performing schools, in spite of NCLBA's intention to narrow the achievement gap (Linn, 2003). Further, NCLBA guidelines seem to expect subgroups to start at the same place and progress at the same rate as the rest of the school, without recognizing the "situation of less advantaged learners" (Abedi, 2004, p.10). For ELLs, their schools have "double duty" (Abedi, 2004), for not only must ELLs learn grade-level content, but they need to become English-proficient in a hurry as well. Thus, AYP requirements and consequences appear to threaten the subgroups NCLBA is designed to help. The requirements set "extremely stringent standards that are highly challenging and possibly quite unrealistic" (Linn, 2003, p. 2).

ELLs as a subgroup

Under the guidelines of NCLBA, a state reports ELLs as one subgroup that either meets or does not meet standard. Several inconsistencies within the ELL group make it a poor candidate for a subgroup: a) students move in and out of this category with regularity; b) the tremendous range within this group cannot be reflected in one subgroup statistic; c) identification procedures for ELLs vary from state to state and within states; d) assessments vary considerably in different states, and e) small groups and large groups

of ELLs are assessed the same, although they do not have the same statistical strength. This section will discuss each of these problems in more detail.

ELLs and mobility

As Washington Public Schools Deputy Superintendent Mary Alice Heuschel explained, "Students move in and out of that group all the time...We'll never be able to measure the success of that group" ("Law: All students", 2003). In 2001-2002, out of the total 72,215 students enrolled in Washington's transitional bilingual program, 20,484 students entered and 14,352 students left (OSPI, 2003a). In other words, almost one half the students enrolled in the program either entered or left the program that year.

New ELL students tend to be low-achieving (Abedi & Dietel, 2004). In Washington, 28.4% of ELLs were new to their district in 2001-2002. The continual entrance of new students exerts downward pressure on the subgroup score; while high end achievers tend to exit out of the program, leaving the subgroup without what balance they might provide (Abedi, 2004; Abedi & Dietel, 2004).

Redesignation. The Council of Chief State School Officers (1992) defines an FEP student as one who is skilled in four classroom areas: a) listening- the student can understand the language of the teacher and instructional discourse; b) speaking- the students can use oral language appropriately in learning activities and in school, including for questioning, testing and challenging of ideas; c) reading- the students can comprehend grade-level text; and d) writing- the student can produce written assignments at grade-appropriate level.

Two major problems result in redesignating ELLs as Fluent English Proficient. The first, mentioned earlier, is that continually removing the higher-end students by

redesignating them as FEP keeps the ELL subgroup very unstable (Abedi, 2004; Abedi & Dietel, 2004). The other is that becoming fluent in English and acquiring English academic fluency is an ongoing process, not measured by the single achievement test often used to determine program exit (Abedi, 2004; CCSSO, 1992; Linquati, 2001).

By continually moving FEPs out of the ELL subgroup, only the lower performing students remain in the program. Keeping FEPs in the ELL subgroup would possibly add some stability to the group. In several studies conducted by the National Center for Research on Evaluation, Standards, and Student Testing (CRESST), a cohort of 14,000 ELL students was followed from ninth through twelfth grade. Each year, achievement of students who exited the language assistance program was compared with achievement of those who remained in the program. Both groups performed below their English speaking peers, but a large gap remained between the FEPs and the students still in program. Also, analyses showed that the group's reading achievement with the FEPs removed underwent a modest but significant drop (Abedi, 2004).

Some states have made keeping FEPs in their ELL subgroup part of their AYP plan. Indiana, Delaware, Georgia and California include FEPs in their ELL group. Recent NCLBA revisions (U.S. Department of Education, 2004) will now allow all states to count their FEPs as part of the ELL subgroup for two years after redesignation.

Exit criteria have typically been norm-referenced tests, normed on the English-speaking population; thus the criteria are often inadequate and in themselves lack validity (Linquati, 2001). Further, states vary in their definition of proficiency; it ranges from the 35th to the 50th percentile (Abedi and Dietel, 2004). As De Avila (1997) noted, the 35th percentile on a state achievement test is not a springboard for academic success. Based on

their studies of achievement of elementary ELLs, both De Avila (1997) and Hakuta et al. (2000) recommended keeping the students in the language assistance program throughout the elementary years, regardless of their test performance. In fact, the percentage of FEPs reading at grade level drops significantly after fourth grade (Linquati, 2001).

The matter of exiting language assistance programs is a "gray area where language proficiency, opportunities for language use, prior schooling, learning ability, academic aptitude, and substantive access to and engagement with academic subject matter all intersect" (Linquati, 2001, p.6). As Thomas and Collier (1997) further demonstrated, since redesignated students meet grade level at minimum levels, they need to make extra gains in order to keep up with their peers. Thus, ELLs who meet exit criteria need both access to language assistance programs after redesignation and regular monitoring (CCSSO, 1992; Gándara & Merino, 1993; Linquati, 2001). The redesignation of ELLs is an issue for more careful study; in the meantime, ELLs should not be exited on the basis of a test (Linquati, 2001).

Wide range of ELLs

The range of ELL educational skills is vast. As LaCelle-Peterson and Rivera (1994, p. 60) stated:

> A newly arrived 11-year-old immigrant ELL, whose schooling has been interrupted by war and frequent moves, and who has not developed literacy in his or her native language, faces relatively greater educational challenges than another recently arrived 11-year-old ELL who has developed literacy skills in his or her native language, succeeded well academically, and already studied some English in school before coming to the United States.

This range extends from a growing number of immigrant students with limited formal schooling (Alcala, 2000; Teachers of English to Speakers of Other Languages, 1997) to U.S.-born students fluent in English language but struggling for English academic fluency to students with high L1 literacy development now seeking L2 language and literacy development. ELLs differ along many dimensions: English language development and educational level upon entry, specific schooling experiences (such as instructional methods and curricular sequences), primary language, country of origin and cultural background, family history and educational background, and socioeconomic status (Collier & Thomas, 2004; CCSSO, 2000; LaCelle-Peterson & Rivera, 1994; McKeon, 1994; Thomas & Collier, 1997).

Studies conducted by the National Center for Research on Evaluation, Standards, and Student Testing (CRESST) showed that within ELL subgroups, performance varied along cultural-linguistic lines *as widely as it did in the general student population.* They found significant differences in achievement between Chinese-speaking ELLs and Spanish-speaking ELLs (Abedi & Dietel, 2004). In another example, both ELL and non-ELL students from a Chinese-speaking background had higher achievement on science and reading tests than students from a Spanish-speaking background (Abedi & Dietel, 2004). Other studies showed parent education level to be a significant factor in ELL achievement, accounting for as much as 15 percentile points (Abedi, Leon & Mirocha, 2003).

In summary, the similarity between ELLs is their development of English language and English academic fluency. The differences between ELLs are as vast as

those between all U.S. students. Regarding them as a monolithic body is inaccurate (Abedi, 2004; Abedi & Dietel, 2004; Lacelle-Peterson & Rivera, 1994).

Inconsistency of ELL identification

Definitions and processes for identifying students in the process of learning English vary from state to state (LaCelle-Peterson & Rivera, 1994) as well as within states (Abedi, 2004; CCSSO, 2000). In many states or districts, initial identification is made by a home language survey completed during student registration. This survey usually asks if the student speaks a language other than English in the home. The validity of this survey is compromised by parent literacy level, citizenship concerns, and the parents wanting equal educational opportunities for their children (Abedi, 2004). Abedi, Lord and Plummer (1997) found a much lower rate of students speaking English at home when the parents reported than when the students did.

After the initial home language survey, students usually take a language proficiency test to qualify for the school's language assistance program (Linquati, 2001). One study reported an average correlation of .223 between scores on a commonly used language proficiency test (Language Assessment Scales) and district LEP classification codes (Abedi, Leon & Mirocha, 2003). As Linquati (2001) points out, students usually qualify for language assistance programs with an oral language test, but their placement, progress and eventual exit are measured by tests of academic fluency.

Wide range of statewide assessments

Since each state has produced its own assessment system, a wide variety of tests exists. Some are norm-referenced while some are criterion-referenced (Abedi, 2004). Iowa has left test choice up to district; most districts choose Iowa-developed norm-

referenced tests, either Iowa Test of Basic Skills (ITBS) or Iowa Tests of Education Development (ITED) (The Princeton Review, 2003). Some states, such as Illinois, have changed tests already, and others (i.e. Michigan) have asked their state legislatures to approve a change to another test (Higgins, as cited in Daggett, 2004). Thus, it is a fluctuating range of assessments.

States have defined their own proficiency levels as well; many have already redefined them by lowering their cut scores or "softening" their standards (Daggett, 2004). When Michigan changed its definition of a proficient school, for example, the number of failing schools went from 1,513 to 216 (The Princeton Review, 2003).

In a study of proficiency levels in 14 states, researchers calibrated cut scores to create a common measurement scale, across grade levels between states and within states (Kingsbury, Olson, Cronin, Hauser & Houser, 2003). They found that proficiency levels in these states were uneven and unequal. Students making cut scores in third grade reading in Montana and California would not necessarily make cut scores in South Carolina, for instance. Montana appeared to have the lowest cut scores: in fourth grade; standards could be met at the equivalent of the 26th percentile; in eighth grade, at the 35th percentile. Wyoming had the highest cut scores; almost 75% of the fourth and eighth grade students in all 14 states would not meet standard in Wyoming. Additionally, some states showed questionable calibration across grades. California, for example, had cut scores at third grade corresponding to the 51st percentile, at sixth grade to the 46th percentile, at eighth grade to the 54th percentile, and at tenth grade to the 14th percentile. The researchers also found that some subjects had higher standards set than others, i.e. mathematics (Kingsbury et al., 2003).

In conclusion, ELLs are being held accountable in the state in which they find themselves for performance on assessments that differ greatly and have limited comparability.

Statistical problems of comparing different sizes of groups

A minimum number of students for a subgroup has not been set by NCBLA (2002), but it requires each state to define a minimum subgroup size that will yield statistically reliable data. Smaller groups are subject to more performance variation; the larger the group, the more statistical reliability the measurement has. Yet numbers of ELLs vary across the U.S., and "in the case of a large number of states and districts, the number of LEP students is not enough for any meaningful analyses" (Abedi, 2004, p.4). For example, Vermont had less than 1,000 ELLs in 2000-2001, while California had 1.5 million (Abedi, 2004). To say that 30% of ELLs in Vermont are proficient does not have the statistical strength as 30% of ELLs in California.

Linn, Baker and Herman (2002) demonstrated that a group size of 10 would have a standard error of difference of 22.4; therefore, they suggested a minimum group size of 25. Others claim a group size of at least 100, with a standard error of difference of 7.1, is necessary for statistical reliability (Hill & DePascale, 2003). Yet if districts must have 100 ELLs per grade level for a dependable growth rate, fewer districts would be able to report ELL data, valuable information from smaller populations would be lost and results from larger districts might be over-generalized (Abedi, 2004). However, smaller subgroups, with questionable statistical reliability, could affect the larger group's ability to make AYP. In summary, flexibility is needed in order to weigh the advantages and disadvantages of particular situations (Abedi, 2004; Linn et al., 2002).

Summary

In conclusion, joining ELLs into a subgroup for the purpose of making conclusions about the success of school systems does not make good sense. It is not a cohesive subgroup for several reasons: a) it is a very mobile group, with as many as half of its members in this state moving in and out within a year's time (OSPI, 2003); b) students are redesignated as fluent English proficient and removed from language assistance programs perhaps prematurely (De Avila, 1997; Hakuta et al. 2000; Linquati, 2001) and by questionable criteria (Abedi, 2004; CCSSO, 1992; Linquati, 2001); c) the range of ELL student characteristics is as wide as the range of the general student population (Abedi, 2004; Abedi & Dietel, 2004; Abedi, Leon & Mirocha, 2003); d) inconsistencies in definitions of ELLs (LaCelle-Peterson & Rivera, 1994) and identification of ELLs within districts and states makes comparability unlikely (Abedi, 2004; Abedi & Dietel, 2004; Abedi, Leon & Mirocha, 2003; Abedi, Lord & Plummer, 1997; Linquati, 2001); e) the substantial differences between state assessments makes comparability impossible (Abedi, 2004; Daggett, 2004; Kingsbury et al., 2004); and f) subgroup sizes create statistical predicaments (Abedi, 2004; Linn, Baker & Herman, 2002).

This section makes the case that the ELL subgroup is extremely complex and contains many subgroups of its own. Policymakers examining ELL achievement need to see it divided and examined from many different angles. In this study of fourth grade ELL test performance, students are divided by their time in district. This limits one source of instability for the ELL subgroup, the movement in and out of the group. Also,

the study examines another source of instability for the subgroup: the achievement of redesignated (FEP) students, whose scores are typically removed from the larger group.

Difficulties of testing in a non-native language

Many concur that results for ELLs on state assessments should be used only tentatively while students are still acquiring English and cross-validated with other information (American Evaluation Association, 2000; August & Hakuta, 1997; Butler & Stevens, 1997; CCSSO, 2000; Hakuta, 2001; Holmes & Duron, 2000; LaCelle-Peterson & Rivera, 1994; Linn, 1998; National Research Council, 2000; Olson & Goldstein, 1997; O'Malley & Valdez Pierce, 1996; Rivera & Stansfield, 1998; Thurlow & Liu, 2001). The following language barrier-related issues arise when state assessments are used with ELLs: students' knowledge and skills are often under-rated; the high degree of English linguistic complexity on content area tests makes them exceptionally difficult; and native language interference presents further obstacles in both test-taking and scoring. A discussion of these issues follows.

Language barrier and ELL content knowledge

On content area assessments, such as science and mathematics, ELLs may understand the content but the language barrier can prevent them from showing their understanding. As Holmes and Duron (2000) stated, "Testing done through English is first and foremost an English language proficiency exam, not a measure of content knowledge" (p.1). Yet consequences are more than under-representation of students' understanding; ELLs are vulnerable to such inappropriate and lasting outcomes as placement in low-performing tracks of students, non-promotion to the next grade, or

failure to graduate based on test performance (Holmes & Duron, 2000; National Research Council, 2000).

The CCSSO (2000) argued that reading and writing should not be "linked inextricably to mastery in content areas other than language arts" (p. 21); ELLs should have the opportunity to demonstrate content skills and knowledge in ways that do not rely on English literacy. In addition, scoring criteria for content areas should focus on content skills rather than English command (CCSSO, 2000; LaCelle-Peterson & Rivera, 1994).

High degree of linguistic complexity

State assessments tend to use complex linguistic features (Abedi, 2004; CCSSO, 2000). These include long words and sentences, low frequency words, words with double meanings, pronouns far removed from their referents, passive voice constructions, prepositional phrases, passive voice constructions, conditional clauses, subordinate clauses, negation, and comparative structures. To ELLs, this adds to the cognitive load and can underestimate content understanding (Abedi, 2004; CCSSO, 2000).

The following three examples indicate that English linguistic complexity on tests can underestimate ELL performance. In a 2001 study, Abedi and Lord found that making minor wording changes in math content test items raised ELL performance. In similar tests with the National Assessment of Educational Progress (NAEP), rewording test items such that math terms and tasks were retained but restated with less linguistic complexity showed higher performance (Abedi & Lord, 2001; Abedi, Lord & Plummer, 1997). In Washington State, item analyses on reading and mathematics sections on state tests at fourth, seventh and tenth grades revealed reduced performance on items with specific characteristics: passive voice constructions, conditional clauses, complex negations,

sentence length and visuals unrelated to the question's focus (McCall, 2004). However, on items with straightforward language and more focused visuals, ELL performance was heightened and closer to native English-speaking peers' performance.

Translated test versions. Title I amendments Section 1111 (b) (3) and (5) to the Elementary and Secondary Acts of 1965 (Improving America's Schools Act, 1994) declare state assessments shall provide for:

> the inclusion of limited English proficient students who shall be assessed, to the extent practicable, in the language and form most likely to yield accurate and reliable information on what such students know and can do, to determine such students' mastery of skills in subjects other than English....(E)ach state plan shall identify the languages other than English that are present in the participating student population and indicate the languages for which yearly student assessments are not available and are needed...and may request assistance from the Secretary if linguistically accessible assessment instruments are needed.

While the section above clearly refers to assessments in students' native languages, very few states offer tests in other languages (CCSSO, 2003). The following examples illustrate how tests in other languages both vary and help learners. In Colorado, an alternate test for elementary ELLs is used: the Spanish Colorado Student Assessment Program (CSAP) test. A parallel test to the English Colorado Student Assessment Program, it was developed and field tested with Spanish speakers (Escamilla, Mahon, Riley-Bernal, & Rutledge, 2003). A 3-year study in Colorado with the Spanish CSAP showed that third and fourth grade ELLs tested in Spanish scored higher than those tested in English (Escamilla et al., 2003). Oregon's state test, on the other hand, has a "side-by-

side" translation of the math and science parts of the test into Spanish or Russian, the state's two main secondary languages ("Dual-language testing", 2004). As a third grader said, "I did most of it in the English. But when I didn't understand the words in English, I read the words in Spanish...That makes it more fair" ("Dual language testing", 2004). Oregon also allows students to use L1 for open-ended written responses for up to 5 years after they enter U.S. schools.

Translated tests are not for all situations, but rather are best utilized when students have a high degree of literacy in the primary language, such as when school programs are in L1, or when students have recently come from their native country with L1 schooling (Lacelle-Peterson & Rivera, 1994; Olson & Goldstein, 1997; Thurlow, Liu, Quest, Thompson, Albus, & Anderson, 1998). Yet in some cases, students' language may be a dialect that differs from the main language of the region, in which case translated versions may not help (Houser, 1995). Depending on their school's instructional model, many ELLs will have received most of their content instruction in English, and as a result, they may be more familiar with content vocabulary and concepts in English (Butler & Stevens, 1997). These students also might not benefit from a native language test version. If, however, they are able to demonstrate content understanding better in their native language, then they should be tested in L1 (Holmes & Duron, 2000). In the case of Oregon, with the side-by-side translations, students may be helped at varying levels of literacy in both languages. The best test will be, in the words of the Title I amendment, "in the language and form most likely to yield accurate and reliable information on what such students know and can do" (Improving America's School Act, 1994).

Accommodations. Simplifying the format and language of the test and utilizing L1 on the test are examples of accommodations. Most states allow test accommodations for ELLs (Rivera & Stansfield, 2000). Rivera and Stansfield (2000), Butler and Stevens (1997) and Shepard, Taylor & Betebenner (1998) have documented the types of accommodations used by many of the different states (seen in Table 2).

Table 2
Potential accommodations for ELLs on statewide tests

Type of accommodation	Examples
Presentation	Translations of test items or test itself Repetition of oral directions Audiotaped test directions or items Added visual supports Extra examples Simplification of directions or test items Use of glossaries, dictionaries or word lists
Response	Student dictation/ use of a scribe Responses in primary language Responding without writing By computer Marking in the test booklet

Table 2 (continued)

Setting	Use of a comfortable, familiar place Small group setting Side-by-side support with a helper
Timing	Extra time More breaks Extra sessions

The accommodations used most were those of setting or timing (Rivera and Stansfield, 1998), although those two actually help the least of the four with the language needs of the students (Menken, 2000). A wide range of accommodations may be best to serve a wide variety of learners (Butler & Stevens, 1997; LaCelle-Peterson & Rivera, 1994; Thurlow & Liu, 2001). A screening process might help match the accommodations to the student (Butler & Stevens, 1997).

With simplification of test items, there has been modest success. Abedi, Lord and Plummer (1997) found success in reducing complexity of math items on a National Assessment of Educational Progress (NAEP) test with students in low and average math classes (including ELLs); however, those in the highest level classes were not helped. In another study with eighth graders and the NAEP math assessment, Abedi, Lord and Hofstetter (1998) found that simplifying math items made a significant difference on achievement for both native English speakers and ELLs. In a study with fourth graders taking a math assessment, ELLs, special education students and regular students

randomly received different versions of a test (a regular version, a simplified version, and a version with a glossary). While there were no significant differences, all groups performed better with the glossary version (Kiplinger, Haug & Abedi, 2000). Albus, Bielinski, Thurlow, and Liu (2001) performed a study with Hmong students, whose primary language does not appear in written form. They found the use of simplified English dictionaries was moderately successful, especially at the intermediate English proficiency level.

Many questions have been raised about accommodations, including the impact they have on results and how they affect comparisons with other students (Holmes & Duron, 2000). The variability among states' use of accommodations also makes it hard to draw broad conclusions (Holmes & Duron, 2000). Butler and Stevens (1997) caution, "Until there is a sufficient research base to support the use of accommodations and to describe appropriate scenarios for use, a healthy skepticism is warranted" (p.2).

Native language interference

Confusion may arise on items where a shared value of meaning is assumed that is not actually present due to native language interference (CCSSO, 2000). One example of this is words or phrases that mean something very different in the student's native language. (Several years ago, Chevrolet manufacturers realized one reason their Nova models might not be selling well in Mexico could be because "no va" in Spanish means "it doesn't go". Subsequently, they rethought their marketing strategy.) Another example is confusion about syntax or word order: something said in L2 doesn't make sense when it is approached with L1 rules. Money systems are another example; symbols representing one dollar to us may represent one peso to someone from Mexico. Number

symbols are not the same for different countries. Other sources of confusion include homophones and less-known synonyms (CCSSO, 2000)

ELL performance will reveal cultural interference as well (CCSSO, 2000). As CCSSO explains, a parade in many other countries is a political event where political leaders present their agenda. Here, they have an entirely different cultural meaning: they are for diversion, they are for the community, children and other groups participate. A further example from CCSSO describes written responses by several ELLs from developing third world countries: when asked to name the elements that define a healthy lifestyle, they mentioned things such as clean air and water, enough food to eat, and food free from toxins (CCSSO, 2000).

Native language interference will have a substantial effect on ELL writing. On open-ended responses this will be especially apparent (CCSSO, 2000). Common manifestations include inventive spellings or spelling influenced by L1; combining of two or more words into one; and omissions of articles, tense markers and plurals. While these errors are developmentally appropriate for ELLs for quite some time, they are not age appropriate and usually weigh heavily in scoring (CCSSO, 2000). Trained scorers for content area tests could ensure that students are not penalized for language interference. One study with transitional fifth graders taking a reading comprehension test demonstrated that scorers with the same language background were better qualified to identify examples of linguistic interference in orthography, phonology and semantics that English-only scorers missed (Beaumont, De Valenzuela & Trumbull, 2002). For example, the verb "gastar" in Spanish means "to spend" or "to waste". A student's written response included this sentence: "I like wasting money on poor people", when the

intended meaning was *spending* money. Bilingual scorers were able to know the intended meaning (Beaumont et al., 2002, p. 223).

Summary

In this section, these issues related to the difficulty of testing in a non-native language were raised: a) Tests measure English language proficiency as well as content knowledge (American Evaluation Association, 2000; August & Hakuta, 1997; Butler & Stevens, 1997; CCSSO, 2000; Hakuta, 2001; Holmes & Duron, 2000; LaCelle-Peterson & Rivera, 1994; Linn, 1998; National Research Council, 2000; Olson & Goldstein, 1997; Rivera & Stansfield, 1998; Thurlow & Liu, 2001). b) Linguistic complexity is extremely high; as a result, state assessments under-represent ELLs content knowledge (CCSSO, 2000; Abedi, 2004). While most states allow accommodations, few states offer translated versions (River & Stansfield, 2000). c) Linguistic interference is developmentally appropriate for ELLs for several years, yet it will count against them in the scoring process (CCSSO, 2000). These issues suggest that time in district might improve test performance, by providing time to break through the language barrier, untangle linguistic complexity and become familiar with testing structures.

Length of time needed to develop academic English fluency

Considerable debate within the research community centers on the acquisition of school literacy in English: how this process occurs and how long it takes. However, most researchers agree it is a lengthy process, requiring from 4 to 10 years (Collier, 1992; Collier & Thomas, 2004; Cummins, 1981; De Avila, 1997; Hakuta, Butler & Witt, 2000; Ramirez, Yuen & Ramey, 1991; Thomas & Collier, 1997, 2002).

This section begins with a discussion of prevailing past and present theories of the acquisition of English academic fluency. Next, research on the length of time required to acquire this literacy will be reviewed. Further, many variables influence the length of time needed for ELLs to perform at grade level in English: a discussion will review research on these variables.

Development of academic fluency in English.

Many studies have shown that oral English proficiency develops relatively quickly while academic English fluency, such as reading and writing skills in the content areas, takes longer to develop (Cummins, 1981, 1984; De Avila, 1997; Hakuta, 1986; Hakuta et al., 2000). Earlier influential work by Cummins (1984) distinguished between two levels of language competency for the English language learner: Basic Interpersonal Communication Skills (BICS) and Cognitive Academic Language Proficiency (CALP). BICS was defined as surface fluency, the language of the street, television, friends and home, or what Hakuta (1986) called "contextualized language". Tangible cues such as gestures, body language, and the two-way nature of conversation provide this context. Cummins (1984) found that BICS required 1 to 3 years for ELLs to develop. Cognitive Academic Language Proficiency, on the other hand, was defined as the language of textbooks and academic areas, or "decontextualized language" (Hakuta, 1986). It relied on knowledge of the language. Cummins claimed BICS were often enough for the less demanding levels of Bloom's Taxonomy (Bloom & Krathwold, 1977): knowledge, comprehension and application. CALP was needed for the more cognitively demanding levels, such as synthesis, analysis, and evaluation (Cummins, 1984). Cummins found that 5 to 7 years or longer were necessary for CALP to develop.

Cummins also developed a well-known Threshold Hypothesis (1979). This stated that when a learner brought a threshold of knowledge in L1 to the classroom, learning in L2 would automatically cross over that threshold as well. However, for the learner whose primary language development had not reached a threshold of desired strength, reaching that threshold in L2 would be more difficult to reach. Cummins (1979) used the term "semilingualism" to refer to learners who had not reached the threshold in either language. Later, Cummins (1981) changed the term to "limited bilingualism".

While Cummins' work illuminated the complexity and lengthy process of language and literacy development in a new language, other researchers have found limitations to his theories. For instance, Baker and Jones (1998) have pointed out the artificial nature of the divisions between BICS and CALP and one threshold to another. In reality, different variables such as task, subject, and age of learner will affect the relative complexity of a language activity and the placement of such a threshold. Further, as Baker and Jones (1998) argue, a student's low cognitive ability in an area may be perceived incorrectly as a threshold barrier. If a bilingual model of education is utilized, the need for a threshold is superfluous: learning content in L1 provides access to the curriculum (Krashen, 1996).

MacSwan (2000) has criticized the Threshold Hypothesis on many fronts. For one, it is a viewpoint of deficiency, in which time in school is needed to teach missing language. As he points out, it is applied only to ELLs. All children continue to add new vocabulary and grammatical refinements into their school years (Chomsky, 1969). Cummins seems to be confusing language proficiency with language, MacSwan claims (2000). Literacy involves vision and motor skills, which have nothing to do with

language. Further, children learn language effortlessly, yet not so with literacy. MacSwan summarizes that literacy achievement is easier to attain when home experiences and schools requirements are closely aligned, but it is not an element of language. Similarly, Gándara and Merino (1993) argued that, although they appear to be correlated, English language development and English academic development are two separate competencies.

What Cummins called 'semilingualism' (1979) and later 'limited bilingualism' (1981) is an erroneous concept; if this were true about a learner, it would mean that they could not acquire their own community's language. This "implies that entire cultures and communities that have rejected literacy and formal schooling have a 'less complex' language than those in the 'literate' world" (MacSwan, Rolstad & Glass, 2002, p. 218). Were this true, it would refute findings from linguists: that all languages possess like mechanisms of complexity for substance and structure (Chomsky, 1965; Pinker, 1994).

MacSwan and Rolstad (2003) suggest replacing the BICS/ CALP theory with a new model, the Second Language Instructional Competence (SLIC). In this model, students first acquire sufficient English to understand all-English academic content. It does not rely on cognitive ability or native language development, nor does it "ascribe any special status to the language of school...(I)t simply denotes the stage of L2 development in which the learner is able to understand instruction and perform grade-level school activities using the L2 alone" (MacSwan & Rolstad, 2003, p. 338).

Length of time research.

Studies have shown that ELLs need 4 to 10 years to perform with their grade level on achievement tests (Collier, 1992; Collier & Thomas, 2004; Cummins, 1981, 1984; De

Avila, 1997; Hakuta et al., 2000; Ramirez et al., 1991; Thomas & Collier, 1997, 2002).

Three large studies performed at the elementary school level concluded that the entire span of years from kindergarten through fifth or sixth grade is crucial for English academic skills development for most ELLs (De Avila, 1997; Hakuta et al., 2000; Ramirez et al., 1991). In Hakuta et al.'s 2000 study, elementary ELL achievement was examined in four school districts: two in California and two in Canada. ELLs from the most successful districts still needed 3 to 5 years to achieve oral proficiency (including oral expression, listening comprehension, use of idioms and complex vocabulary). Academic proficiency in English reading and writing took 4 to 7 years, if students began at a school in kindergarten and continued there without interruption. Students who moved or entered later needed longer.

However, Rossell (2000) criticized the test design and conclusions of Hakuta et al. (2000). The tests used to measure English academic fluency also measure academic ability, she argued. Therefore, it was erroneous to draw conclusions about degree of English acquisition.

De Avila's 1997 study investigated several thousand ELL student records. He documented growth in students' first 2 years of listening and speaking achievement, with very little growth in reading and writing. In the third year, after students had reached a level of conversational fluency, their reading achievement began to rise. In the fourth year, their writing achievement began to rise. He found that it took about 5 to 6 years for new ELLs to meet a minimum academic proficiency level of 30 to 40% of expected grade level performance. This happened to be the exit criteria for "graduating" from the bilingual program in the students' state, California. Yet De Avila demonstrated that their

achievement had not been in motion for long at the exit junction. He concluded that at this point, students being redesignated as fluent were in danger of academic failure.

In a Congress-mandated study, Ramirez et al. (1991) followed 2,000 elementary students for 4 years in three different program models- structured English immersion, transitional bilingual and maintenance bilingual. While Ramirez and the other researchers found no differences in program effectiveness over that length of time, they did find that 4 years of support were not enough. One of their recommendations was to give ELLs at least 6 years of program support.

Collier and Thomas (Collier, 1992; Collier & Thomas, 2004; Thomas & Collier, 1997, 2002) performed several length-of-time studies, with similar findings. Their 1997 longitudinal study utilized more than 700,000 ELL student records, for a period of 12 years (from 1982 to 1996), in five large U.S. school systems. Students came from 150 different home language backgrounds. They found that students required from 4 to 10 years to reach English academic fluency. In a recent 2004 study, they examined the largest set of databases yet in ELL educational research: more than 2 million student records, from 23 school districts in 15 states, over a time-span of 18 years. Records came from a cross-section of districts- rural, urban, low-income, high-income, different geographic regions. They found that it took longer than they originally claimed: it consistently required at least 6 to 8 years to close the gap. However, Thomas and Collier (1997) identify academic proficiency as the 50th percentile on achievement tests; states' definition of proficiency ranges from the 35th or 40th to the 50th percentile (Abedi & Dietel, 2004).

Variables influencing length of time required.

Researchers identify several variables that influence the length of time ELLs need to develop academic readiness. While researchers debate the relative strength of these variables, these are most commonly identified: native language development; socioeconomics, educational background and literacy development in the native language; educational level of parents; school program model; and quality of instruction (Collier, 1991; Collier & Thomas, 2004; Cummins, 1981, 1984; De Avila, 1997; Hakuta et al., 2000; Ramirez et al., 1991; Rivera & Stansfield, 2000; Thomas and Collier, 1997, 2002). In this section, I will summarize main research for each variable.

Native language development. The degree of native language development has been identified as a strong predictor for success in second language development (August & Hakuta, 1997; Collier, 1991; Collier & Thomas, 2004; Cummins, 1979, 1981, 1984; Hakuta, 1986; Ramirez, et al., 1991; Thomas & Collier, 1997; 2002). As previously discussed, Cummins' (1981, 1984) found that strong language development in a native language transferred to a second or subsequent language and proposed a Threshold Hypothesis (1979), where L1 academic language development would allow ease of L2 academic language development. Collier & Thomas (2004) claim that students born in the U.S. who acquire both the primary language and English before kindergarten meet mixed success. When placed solely in a regular English-speaking classroom, bilingual needs of these students are sometimes not met, resulting in a seeming failure to flourish and an inaccurate label such as "low language" (Brisk, 1998; Collier & Thomas, 2004).

Educational background and literacy development in first language. Research indicates that having some formal schooling and literacy development in L1 makes

acquisition of language and literacy development in L2 easier (Alcala, 2000; Berman, 1997; Brisk, 1998; Collier & Thomas, 2004; Cummins, 1986, TESOL, 1997; Thomas & Collier, 1997). In their study, Thomas and Collier (2002) found degree of native language schooling to be *the strongest predictor* of academic success in English. They identified the ages of 8 to 11 as the optimal age of arrival in U.S. schools. If the student had attended school until then, acquiring literacy in the primary language first, it would transfer along with language development. They would not have the added burden of learning to read and write from the beginning, on top of learning a new language. And they would have enough years left in school to be able to catch up with their peers, while newcomers arriving after age 12 did not usually have enough years to catch up.

Berman (1997) also found that that newly arrived ELLs under age 12 with at least 2 years of schooling in their native language could reach peer achievement levels in 5 to 7 years, but those without 2 years or more could take as many as 10 years. Similarly, Cummins' theorized from his Threshold Hypothesis work that a student who had schooling in their native language until age 10 could reach academic English fluency more readily (1986).

However, with the increasing numbers of ELLs enrolling in U.S. schools, there have also been increasing numbers of older students (age 12 and up) with two or less years of schooling in their home country (Alcala, 2000; Brisk, 1998; TESOL, 1997). Reasons for the lack of schooling may include living in a rural area where schooling was not available, having to work to support their family or help at home rather than attend school, moving frequently or living in a war area (TESOL, 1997). They might be

preliterate (Alcala, 2000). Most will not be able to close the academic gap during the school years (Collier & Thomas, 2004).

Educational level of parents. The amount of formal schooling of parents has been found to be a strong variable in rate of ELL achievement of English academic fluency (Abedi & Dietel, 2004; Abedi, Leon & Mirocha, 2003; Collier & Thomas, 2004; Hakuta et al., 2000; Thomas & Collier, 1997). Hakuta et al. divided students' parents from their 2000 study into educational levels completed: they showed that students with parents whose education went beyond high school were considerably ahead of other students. Similarly, Abedi and Dietel (2004) found 15 percentile points between reading scores of ELL students whose parents had postgraduate educations and those whose parents did not. In another study, Abedi, Leon and Mirocha (2003) found that ELLs with parents with postgraduate educations had higher reading scores than *non-ELLs* whose parents did not have postgraduate schooling.

School program model. The research community as well as the public has debated school program models at great length. Effectiveness of different program models for ELLs has been the subject of many studies, with slightly more studies favoring a long-term bilingual model such as maintenance bilingual or dual language programs (Baker & de Kanter, 1981, 1983; Collier & Thomas, 2004; Greene, 1998; Ramirez et al., 1991; Rossell & Baker, 1996; Thomas & Collier, 1997, 2002; Willig, 1985). School program appeared as equally strong and sometimes stronger a variable than socioeconomic status in effect on length of time to reach English academic fluency in Thomas and Collier's studies (1997, 2002, 2004). They found that students from the programs with first language (L1) development sustained better achievement over several years than English-

only programs, as measured by standardized reading tests (Terra Nova, Stanford 9, ITBS, CTBS, SABE and Aprenda 2). ESL programs were substantially weaker than the bilingual program models. Transitional bilingual (transitioning to English only within 3 years) showed the least of the bilingual program gains but more than the English-only gains. Maintenance bilingual (transitioning to English only within 5 to 6 years) programs showed better success, but dual language programs showed the most gains of all. In dual language programs, monolingual English as well as non-English speaking students together study in both English and the native language (L1) of the non-English speaking group, perhaps continually throughout the school years. Collier and Thomas (2004) found achievement for this group eventually passed the 50th percentile, after 6 to 8 years.

An early review by Baker and de Kanter (1981, 1983) of 39 studies of various U.S. and Canadian language assistance programs led the researchers to find English-only and transitional bilingual programs as superior. Willig (1985), however, used a statistical meta-analysis approach to examine 23 studies from Baker and de Kanter's review. She chose U.S. studies only, excluding Canadian immersion programs; her findings differed from Baker and de Kanter's. She found that bilingual programs tended to produced superior results on achievement in reading, language arts and mathematics. Further, she found that the more well-designed the study was, the stronger the gains were. Another finding was the wide variety within the category of bilingual programs, making generalization difficult (Willig, 1985).

Rossell and Baker (1996) carefully selected 72 studies comparing transitional bilingual programs to other programs. Their review found that transitional bilingual programs produced slightly better gains than ESL and submersion (sink or swim)

programs, but structured immersion programs outperformed transitional bilingual programs. However, they were criticized for measuring effects of bilingual programs after just one year.

In another meta-analysis, Greene (1998) selected 11 studies from the Rossell and Baker (1996) study. While he believed their criteria for study selection to be sound, he found that relatively few of Rossell and Baker's studies actually met their own criteria for inclusion in the study. Greene showed that ELLs using at least some of the native language for instruction outperformed ELLs that did not (Greene, 1998).

As previously mentioned, Ramirez et al. (1991) followed progress of 2,000 elementary students in three different program models for 4 years to measure model effectiveness. The models they studied were a structured English immersion program, an early-exit or transitional bilingual program, and a late-exit or maintenance bilingual program. The researchers found no differences in achievement between programs, few differences between instructional strategies, but students in all three models were behind their English-speaking peers (as measured by the CTBS). However, criticisms of the Ramirez study noted that 4 years were not long enough to see enough academic growth in any of the programs (Baker & Jones, 1998).

However, McLaughlin and McLeod (1996) found that successful schools were flexible in program construction: rather than using one type of program, they adjusted curriculum and language of instruction to meet the needs of the student. August and Hakuta (1997) similarly recommended that in place of identifying one best program to rather identify a set of components that work best. Certain programs might work best in certain settings, depending on such variables as common goals, resources, and

demographics. Miramontes, Nadeau and Commins (1997) and Brisk (1998) also stress that a "models" approach is inadequate; congruence between theory and practice is more crucial. They found a strong school-wide commitment to be a critical component.

In many schools and districts, bilingual programs are not feasible due to small numbers of language groups. In the states of Arizona, California and Massachussets, native language instruction has been banned by English Only referendums. In 54% of the nation's schools, programs for ELLs use solely English for the language of instruction (Kindler, 2002). Instruction in a native language is more common at the elementary level than the secondary level (Kindler). When the second language is used, it is usually Spanish; 77% of ELLs in the U.S. are Spanish speaking (U.S. Department of Education, 2004).

In Washington State, where this study was conducted, a study of 46 school districts found that students in programs with L1 instruction were able to meet exit criteria (achievement at the 35th percentile on standardized achievement tests) from the state bilingual program more quickly than the students in ESL programs (OSPI, 2000c).

Quality of instruction. When students could be followed longitudinally in schools where staff received intensive training in interactive teaching approaches, student achievement both improved and held (Thomas & Collier, 1997). These approaches included literacy development across the curriculum, process writing, thematic lessons, cooperative learning, portfolio assessment, and uses of technology.

Researchers agree that effective schooling for ELLs occur in buildings where their education is reflected as a school-wide endeavor and where teachers and staff are adequately prepared to meet ELL learning needs (Berman, 1992; Brisk, 1998;

Miramontes et al., 1997; Ramirez et al., 1991; Thomas & Collier, 1997). This condition appears to be the exception rather than the norm. Moss and Puma (1995) found that 58% of ELLs in high poverty schools received their English and reading instruction from aides; nationally, only 21% of aides who work with ELLs have post-high school education. In a survey of 422 teachers in the Midwestern U.S., Walker, Shafer and Iiamo (2004) reported that 87% of teachers working in schools with incoming ELLs had no previous professional development for ELL students. This corresponds closely to McCloskey's 2002 findings (as cited in Walker, Shafer and Iiamo, 2004) that 88% of U.S. teachers have no professional development for working with English language learners, while an estimated 43% have ELLs in their classrooms (U.S. Department of Education, 2004). Most teacher certification programs do not require courses in working with ELLs (Crawford, 1993).

The American Educational Research Association (2004) summarizes from research the best ways to teach ELLs: in small groups with one-on-one tutoring; with the best teachers available; with lengthened school days and extended school years; by providing several years of intensive high-quality instruction for academic fluency; and by extensive supported, explicit practice reading with ongoing assessment.

Researchers indicate the practices that follow will build ELL success. Cultures and home languages of the students must be viewed as resources; L1 should be used as a foundation as much as possible, and family participation at school should be sought (Berman, 1992; Brisk, 1998; LaCelle-Peterson & Rivera, 1994; Wong-Fillmore, 1990). Active learning environments, use of thematic units and learning groups, language-rich environments, successful integration of language and literacy, alignment of assessment

with curriculum are necessary to make content comprehensible and accessible (Brisk, 1998; Miramontes et al. 1997; Thomas & Collier, 1997). Adequately prepared teachers with clear goals and high expectations need to keep a continual emphasis on language development and responsiveness to individual needs of students (August, Hakuta & Pompa, 1997; Brisk, 1998). Teachers trained in content delivery and language sheltering that continually builds academic fluency are shown to effect student achievement (Short & Echeverria, 1999; Solomon and Rhodes, 1995).

Socioeconomics. In their 2000 study of four school districts in California, Hakuta et al. found that ELLs from lower socioeconomic status (SES) learned English more slowly on the average: "(i)t now appears certain that SES is powerful in predicting rate of English acquisition" (p. 13). Krashen (1996) called socioeconomic status "defacto bilingual education", meaning that parents of higher socioeconomic groups gave their children a language-building upbringing that aided their academic English fluency. The study by Ramirez et al. (1991) identified socioeconomic status as a decisive factor in ELL student achievement.

Thomas & Collier (1997), however, found socioeconomic status harder to measure with ELLs. In their 1997 study of more than 700,000 student records, more than 57% qualified for free and reduced lunch. Yet because most were immigrants, there were unanswered questions: what were the families' socioeconomic backgrounds in their home countries? Many may have been middle class or above. What were the educational levels of the parents? What are the family aspirations for the future? Thomas and Collier did not find SES to be one of the strongest predictors of ELL achievement. As mentioned, besides amount of formal schooling in the first language, they found type of

school program to be more predictive of school success than SES. Particularly, students enrolled in dual-language bilingual programs made the most gains.

Rolstad (as cited in MacSwan, 2000) suggested the term *sociolinguistic status*. This combination of socioeconomic status and linguistic status carries a corresponding degree of empowerment, so that an ELL might feel at the bottom of the heap or relatively accepted in middle class. The problem lies in feeling at the bottom of the sociolinguistic ladder, which can lead to a cycle of poor performance and low expectations.

Summary

It is difficult to isolate these variables to demonstrate how each effects rate of acquisition of academic English fluency. The variables are often interrelated, as the following examples show. Lower socioeconomic status often means living in a low rent area where schools have higher percentages of low income students; as Moss and Puma (1995) found, 58% of ELLs in high poverty schools received their English and reading instruction from aides. Thus, lower SES often goes with lower quality educational programs. Thomas and Collier (1997) found that formal schooling in native language before arrival in U.S. often was accompanied by higher parent education levels. (When this was the case, it was an extremely powerful predictor of ELL achievement.)

Assumptions about variables can also be misleading. While type of school program is a strong predictor, schools vary considerably in implementation. A particular school with a maintenance bilingual model can be of weak quality and therefore lose its strength as a predictor of achievement; a school with an ESL program can have high achievement (Brisk, 1998). Parents in a low SES home in the United States can be from an upper middle class home in their country of origin (Thomas & Collier, 1997).

In summary, examining these variables identified by researchers reveals the complexity of the process of acquiring English academic language. It helps us predict circumstances that may make the process slower or faster. Yet they are hard to isolate and define precisely and exclusively. To examine a large body of students, time in district appears more functional.

Suitability of statewide assessments for ELLs

This section will begin with a discussion of suitability of high stakes assessments for ELLs. Next, development of the Washington Assessment of Student Learning and its suitability for ELLs will be addressed. Achievement of the fourth grade ELL subgroup on the WASL for the 3 school years of this study will be then be reviewed. Finally, the questions for the study will be presented.

High stakes assessments and ELLs

This statement by CCSSO (2000) is echoed by many teachers and administrators: "Large scale assessments are studded with problems that affect not only how equitably the achievement of LEP students can be measured, but also how effectively LEP students' mastery of content is assessed" (p.5). Various studies have surveyed educators on their beliefs toward their states' achievement tests as a measurement of ELL performance. Findings show that validity is questioned (Mabry, Poole, Redmond, & Schultz, 2003; Moon, Callahan & Tomlinson, 2003), that testing is regarded as biased and not fair for ELLs (Mabry et al., 2003; Olson, 2002; Schulte, Schulte, Slate & Brooks, 2002; Wright, 2002), that test results are not useful (Mabry et al., 2003; Wright, 2002) and that teachers have narrowed the curriculum for more test preparation, resulting in less emphasis on higher thinking skills (Mabry et al., 2003; Moon et al., 2003; Olson, 2002;

Wright, 2002). Generally, educators prefer multiple measures for ELL assessment (Mabry et al., 2003; Moon et al., 2003; Olson, 2001; Wright, 2002).

As previously reviewed, many researchers question the appropriateness of high stakes assessments for ELLs on grounds that it measures both English proficiency and content; further, they contend that assessments should be cross-validated with other measures (American Evaluation Association, 2000; August & Hakuta, 1997; Butler & Stevens, 1997; CCSSO, 2000; Hakuta, 2001; Holmes & Duron, 2000; Lacelle-Peterson & Rivera, 1994; Linn, 1998; National Research Council, 2000; Olson & Goldstein, 1997; O'Malley & Valdez Pierce, 1996; Rivera & Stansfield,1998; Thurlow & Liu, 2001).

Rivera and Stansfield (1994) identify two undesirable practices used in the past for inclusion of ELLs in large-scale testing: including ELLs as though their scores meant the same thing as native English-speaking peers and exempting them altogether. Warning against the same practices in the present, Rivera and Stansfield (1994) advocate an assessment program that tracks individual students over time and reports their increasing ability in English language and content. Also, they suggest accountability assessments can become a worthwhile opportunity for further meaningful disaggregation of student background information (Rivera & Stansfield, 1994).

The position of the Council of Chief State School Officers (2000) is that "LEP students should be included equitably whenever possible in large-scale assessment" (p.6). Through a series of measures instituted in the psychometric community over the last 20 years and described below, the CCSSO (2000) addresses steps to make assessments more equitable: they include test developing and piloting that includes a full range of ELLs,

item reviews for both bias and differential item functioning (DIF), and specific scoring procedures.

Test development and piloting. Most state assessments were developed for native English speakers and field-tested with native English-speakers (Abedi, 2004; CCSSO, 2000; LaCelle-Peterson & Rivera, 1994; Menken, 2000). As a result, the tests assume English language proficiency. Yet for ELLs, the tests measure language proficiency as well as content knowledge: test items require both linguistic processing and cognitive processing (August & Hakuta, 1997; August et al., 1994; Hakuta, 2001; Holmes & Duron, 2000; LaCelle-Peterson & Rivera, 1994; Menken, 2000; Olson & Goldstein, 1997; Rivera & Stansfield, 1998; Thompson, DiCerbo, Mahoney & MacSwan, 2002). Thus, tests lack content validity where ELLs are concerned, because they measure other than that which they claim to measure (O'Malley & Valdez Pierce, 1996). Further, when validity studies have not been performed on tests for ELL students, students' language proficiency status is a likely source of measurement error (Thompson et al., 2002).

Test developers, both states and publishers, should include samples representing the full range of their ELL population in the field testing process (CCSSO, 2000). This includes all combinations of languages and proficiency levels within the testing area. Only in this way can assessment be aligned with all of its test population. Further, content area tests should allow a variety of performances, some non-text and non-English, to provide students alternate means of demonstrating knowledge. Examples include oral explanations, use of charts and pictures (CCSSO, 2000). Additionally, ELL specialists should be involved in assessment technology development and assessment administration (Rivera & Stanfield, 1994).

Item reviews. Test publishers have routine procedures which flag readability problems in test problems, yet these are not sufficient for ELL test-takers (CCSSO, 2000). Bias reviews that examine unnecessary linguistic complexity of items on content area tests would improve suitability for ELLs (Abedi & Dietel, 2004; CCSSO, 2000). Unfortunately, the focus of most state bias committees that examine state assessments has been avoidance of stereotypical language and cultural irrelevancy; a more expanded focus on language bias is needed (CCSSO, 2000). Bias committees often consist of citizens and educators; for language bias reviews, linguists, bilingual specialists and other experts should be included as well (CCSSO, 2000).

Additionally, differential item functioning (DIF) procedures should be carried out by test companies and/or state education agencies periodically to identify differential performance by ELLs and other student subgroups (CCSSO, 2000; Shepard, Taylor, & Betebenner, 1998). This procedure flags items that are particularly troublesome and allows test publishers to make the items more accessible. By examining ELL performance on WASL test items for 2002-2003, McCall (2004) found different achievement patterns on distinct item types; on items with plain language and clearly related visuals, ELLs performed noticeably higher than on items with complex phrasing, extended sentence formations and semi-related visuals.

Scoring. Scoring on content area open-ended items should not penalize students for lack of English (LaCelle-Peterson & Rivera, 1994).Thus, scorers should be specifically trained to score ELL work (CCSSO, 2000). Training should include recognition of characteristics of language development and native language interference (Butler & Stevens, 1997; CCSSO, 2000). Rubrics can be especially developed for ELLs;

unless literacy has been chosen for evaluation, it should not be confounded with content measure evaluations (CCSSO, 2000; McKeon, 1994).

The CCSSO acknowledges that its measures have not been utilized widely and that there remains "widespread misunderstanding (in the testing and education communities) about what exactly mainstream tests actually measure relevant to the achievement of LEP students" (p. 14). However, their demonstrations (for example, of ways to reduce linguistic complexity in a passage and visual clutter on a page) provide a model to straightforward model to follow.

Summary. Teachers, administrators and researchers share in skepticism about statewide tests as a valid and fair measure of ELL achievement (American Evaluation Association, 2000; August & Hakuta, 1997; Butler & Stevens, 1997; CCSSO, 2000; Hakuta, 2001; Holmes & Duron, 2000; LaCelle-Peterson & Rivera, 1994; Linn, 1998; Mabry, Poole, Redmond, & Schultz, 2003; Moon, Callahan & Tomlinson, 2003; National Research Council, 2000; Olson, 2002; Olson & Goldstein, 1997; O'Malley & Valdez Pierce, 1996; Rivera & Stansfield,1998; Schulte, Schulte, Slate & Brooks, 2002; Thurlow & Liu, 2001; Wright, 2002). Results should be cross-validated with other sources of data (American Evaluation Association, 2000; August & Hakuta, 1997; Butler & Stevens, 1997; CCSSO, 2000; Hakuta, 2001; Holmes & Duron, 2000; LaCelle-Peterson & Rivera, 1994; Linn, 1998; National Research Council, 2000; Olson & Goldstein, 1997; O'Malley & Valdez Pierce, 1996; Rivera & Stansfield,1998; Thurlow & Liu, 2001). High stakes tests might become more equitable if measures suggested by the Council of Chief State School Officers (2000) were followed, such as including the complete range of the area's ELL population in test development and field testing, conducting item bias reviews and

differential item functioning reviews, and training scorers specifically on characteristics of English language development and linguistic interference.

Assessment of ELLs in Washington State

ELL education. Language assistance programs exist in 63% of Washington's schools; in 2000-2001, programs served 70,431 students with a total of 181 different home languages (OSPI, 2002c). Spanish speakers account for 62% of Washington's ELLs. In some districts, this group of students has grown at a phenomenal rate, as in the town of Wenatchee, where Spanish-speaking students have increased by 551% in 12 years (Sanchez, 2000). Another 24% of Washington's ELL population is made up of Russian, Ukrainian, Vietnamese, Korean, Cambodian, Somali, and Tagalog speakers combined (OSPI, 2002c). Twenty-one districts serve more than 20 different language groups; Kent and Seattle School Districts each serve more than 60 different language groups (OSPI, 2002c).

While the state utilizes the label "transitional bilingual instruction program", schools implement either a bilingual or English as a Second Language program models. A 3-year exit is the state's goal; 28% of the students in the state bilingual program have been there for more than 3 years (OSPI, 2002c). At the other end, in 1999-2000, 31% of the states' ELLs were newly enrolled (OSPI, 2000a). Exit criteria was the 35th percentile or above on a norm-referenced achievement test such as the ITBS (OSPI, 2002c), but under NCLBA guidelines the Washington Language Proficiency Test Series (MetriTech, Inc., 1999) is now used for both annual measurement of language development and exit determination. The test (WLPT) has a mixed format, with some multiple choice and some open-ended written response items.

Washington Assessment of Student Learning. In Washington, the assessment system is based on standards that reflect broad learning goals, the Essential Academic Learning Requirements, or EALRs. Content committees in each curricular area (originally reading, writing, math and listening) made up of about 20-25 citizens and educators in each area, created a pool of items for tests for fourth, seventh and tenth grades (Taylor & Lee, 2001). A fairness review committee checked for appropriate difficulty, fair and culturally unbiased content, appropriate guidelines; also, they made sure that there was only one best answer for each test item (Taylor & Lee, 2001). *However, there were no reviews for linguistic complexity* (McCall, 2004). These tests are criterion-referenced; they include various kinds of items- multiple choice and short answer questions as well as essays and problem-solving tasks.

Fourth grade tests were piloted with 85% of the state's fourth graders in spring 1996. Performances of categorical groups, such as bilingual program students and Title I students, were noted but not analyzed (Taylor & Lee, 2001). Standards were set in spring 1997, and the first actual test was given in spring 1998.

In determining the cut-scores for each level, the following statement was used as a guide: "In all content areas, 'meets standard' reflects what a well-taught, hard-working student should know and be able to do" (Taylor & Lee, 2001, 6-2). In the reading and math tests, there are four possible scores: levels 1 (well below standard), 2 (below standard), 3 (meets standard) and 4 (exceeds standard). For the writing and listening tests, there are simply two scores, below standard and meets standard.

Brickell and Lyon (2003) reviewed 17 different studies and technical reports on the WASL and found several problem areas; among them was that WASL stratified along

ethnic lines, with Caucasian and Asian/ Pacific Islander students having the highest passing rates and Hispanics having the lowest. Also, reading and math WASL tests had higher than expected correlations, indicating that the math test was a measure of math and reading (Brickell & Lyon, 2003).

Accommodations for all students on the WASL are allowed in the areas of scheduling (i.e. providing more testing time and breaks, except for on the writing test); setting (providing study carrels, private space, preferential seating, small groups or familiar environments); presentation (including administering directions in primary language, or rereading or tape recording of directions, providing side-by-side assistance in handling materials and staying on task; allowing students to use dictionaries on writing tests); and response (providing a scribe). For ELLs, if they fall within the "limited English speaker range" (levels 1 through 3 on the state oral language proficiency test), students may have two extra presentation accommodations: using a reader to read mathematics or science assessment prompts in English, verbatim; and using a dictionary and/ or thesaurus in English, native language, or visual formats during the writing assessment (OSPI, 2003b).

Under Washington's AYP plan, by 2008, 80% of all elementary students must meet or exceed the standards. Also a certificate of mastery will be given to students who meet the tenth grade standards; starting in 2008, this certificate will be necessary for high school graduation (Commission on Student Learning, 1998). In 1999, norm-referenced tests for third, ninth and eleventh graders (later changed to third, sixth and ninth graders) were also added (OSPI, 2000b).

ELL achievement on WASL. WASL scores for state fourth grade students in the 3 academic years of this study show gains, although scores are far below the 80% needed by 2008 (as shown in Table 3). (However, in a 2003 study, Orlich applied statistical analyses of effect sizes to the testing years of 1998-1999, 1999-2000, 2000-2001 and found that there actually were no gains.) State ELL performance on the WASL also appeared to rise slightly, but scores remain considerably lower than those of their native English-speaking peers (also in Table 3). Also, the gap seen here between ELLs and native English-speaking peers is often larger than the 20 to 30 % average gap found by Abedi (2004) between ELLs and monolingual peers on achievement tests.

Table 3
Percentage of All Fourth Grade Students and ELL Students Meeting WASL Standards for 2000 –2002, Washington State

	1999-2000		2000-2001		2001-2002	
Test	All	ELL	All	ELL	All	ELL
math	41.8	10.9	43.4	11.6	51.8	18.2
total n=	75,977	2,215	75,492	4,151	75,282	4,491
reading	65.8	20.9	66.1	24.0	65.6	24.8
total n=	75,733	2,191	75,195	4,129	75,074	4,468

Table 3 (continued)

listening total n=	65.3 76,125	30.0 2,217	72.4 75,905	36.5 4,177	66.6 75,761	35.4 4,524
writing total n=	39.4 74,883	10.5 2,113	43.3 74,296	15.4 3,969	49.5 74,358	19.4 4,319

Note. From Office of Superintendent of Public Instruction, 2002b [On-line], WASL test scores. Available from http://www.k12.wa.us/edprofile/

State Superintendent Bergeson and Deputy Superintendent Heuschel (2003) issued a proposal requesting flexibility in three areas for determination of AYP for Washington State. First, they requested that individual states be allowed to petition the Department of Education "to present a quality plan that includes alternative strategies to accomplish the mission and goals of NCLB, as long as the alternative state plans address the major elements within the law and reflect the goal of 100% proficiency by 2014" (p.1). Second, they proposed that students in Special Education and Bilingual/ LEP programs, already in the accountability system within their specialized education program, be allowed "a more constructive and valid accountability approach" to assessment (p.1). Third, they propose changes in the accountability mechanisms; for example, states would have more flexibility for continuous growth models, for determining AYP baselines from schools' and districts' current level of performance, with states setting increments for improvement (Bergeson & Heuschel, 2003).

Summary

In Washington, the state assessment development process did not specifically

address needs of ELLs. Test development and field testing did not include ELLs, bias committee screenings do not examine items for linguistic complexity, there have been no differential item functioning procedures, scorers have received no special training for ELL tests. Results for fourth grade ELLs for test years 1999-2000, 2000-2001, and 2001-2002 reveal a large gap. The math assessment appears to confound math achievement with reading (Brickell & Lyons, 2003). However, the State Superintendent and Deputy Superintendent (2003) have issued a proposal suggesting alternative ways to figure AYP that is more fair for ELLs.

As noted in earlier sections of this chapter, putting all ELLs into one subgroup is problematic. In 2001-2002, half of Washington's ELL population either entered or left the language assistance program that year ("Law: All students", 2003). Further, the wide range of variables found within the ELL subgroup make it is diverse as the general student population (Abedi & Dietel, 2004; Abedi, Leon & Mirocha, 2003). The problems of testing ELLs in English when these students are in the process of acquiring English language and academic fluency are many, but perhaps most strongly researchers question the appropriateness and validity of testing all ELLs and using the results for high stakes decisions (American Evaluation Association, 2000; August & Hakuta, 1997; Butler & Stevens, 1997; CCSSO, 2000; Hakuta, 2001; Holmes & Duron, 2000; LaCelle-Peterson & Rivera, 1994; Linn, 1998; National Research Council, 2000; O'Malley & Valdez Pierce, 1996; Olson & Goldstein, 1997; Rivera & Stansfield, 1998; Thurlow & Liu, 2001). In spite of a wealth of research documenting that several years are necessary to reach levels of English academic fluency (Collier, 1992; Collier & Thomas, 2004; Cummins, 1981; De Avila, 1997; Hakuta, Butler & Witt, 2000; Ramirez et al., 1991;

Thomas & Collier, 1997, 2002), policymakers plan for ELLs to take state assessments and achieve at the same levels as native English-speakers.

Statement of the problem

While the fourth grade state WASL scores have shown modest gains, they have to close a sizeable gap in order to reach 80% by 2008. Further, for the 2001-2002 school year, only 36 of 1162 elementary schools in Washington made AYP (Mabry et al., 2003). In a press release, State Superintendent Bergeson (2003) said:

> I have substantive concerns about the manner in which the achievement of our students enrolled in special education and bilingual programs is measured under No Child Left Behind. We need to be held accountable for making academic progress with the children in these two programs, but that progress needs to be measured in a way that is much more fair to these children.

At what point do ELL scores show meaningful participation? Setting time in the U.S. as a criterion is no perfect answer; as Rivera and Stansfield (1998) note, some students display English academic readiness for a state assessment after a short amount of time in the U.S., and others are not ready after a longer time. Many factors affect the length of time needed by ELLs; it is indeed a complex issue (Collier, 1991; Collier & Thomas, 2004; Cummins, 1981, 1984; De Avila, 1997; Hakuta et al., 2000; Ramirez et al., 1991; Rivera & Stansfield, 2000; Thomas and Collier, 1997, 2002).

However, time in district is one area for further investigation. Hakuta, Butler and Witt (2000) reported findings that displayed English academic fluency development as a function of time. They consistently found 4 to 7 years necessary for grade level academic proficiency to develop. De Avila's research (1997) also showed that for several thousand

beginning English language learners in California elementary schools, grade-level competency was attained after an average of 5 years. Earlier large well-known studies by Cummins (1981) found that an average of 5 to 7 years were necessary; Thomas and Collier (1997) similarly found that 5 to 7 years were usually necessary for development of English academic fluency. (In Collier & Thomas, 2004, this was amended to 6 to 8 years.)

These previous studies measured achievement with norm-referenced tests. Their predictive ability for criterion-referenced tests, such as the WASL, cannot be assumed. However, Joireman and Abbott (2001) performed analyses on more than 45,000 fourth graders' WASL and Iowa Test of Basic Skills scores; they found that if a student scored high on one, they were likely to score high on the other. They reported that tests measured similar but not identical abilities, especially in the domains of reading and math. Taylor (1998) similarly found that the constructs measured by the WASL and California Test of Basic Skills were interrelated to a moderately high degree.

The purpose of this study was to investigate a possible relationship between time in district for fourth grade English language learners in five districts in southeastern Washington and their performance on the WASL. The data from this study may inform key administrators about how ELL scores change with time and at what point they begin to approach the scores of their native English-speaking peers. This study addresses one piece of a complex puzzle; by itself, time in district cannot predict rate of ELL achievement, Yet it remains a central factor. Presentation of this data may aid understanding and help shape district practice and policy.

The study will attempt to answer the following questions:

1. How does WASL test performance for ELLs differ at 1, 2, 3, 4 and 5 years in district?
2. At what point do ELL scores become similar to the scores of their native English-speaking peers?
3. How do test data for students who have exited the bilingual program compare to those for students still in the program and with other mainstream students?
4. If the districts' ELL scores, grouped together and unreflective of different lengths of time in district, are reported separately from the mainstream, how do the mainstream scores change?
5. What will administrators from the five districts think about the data?
6. After reviewing the data, what do administrators from the five districts recommend as an equitable way to include ELLs in the state assessment system?

Chapter 3

Methodology

This chapter begins by briefly describing the study design and situating the research questions within the design. It then addresses the study sample and setting, with descriptions of the five neighboring school districts that participated in the study. Following that, the strategies utilized in the study for data collection and data analysis are described.

Overview of design

To answer the questions posed by this study, both quantitative and qualitative components were conducted. The quantitative component combined data from five neighboring school districts to provide a larger sample. To describe a relationship between length of time in district and WASL achievement of fourth grade English language learners, WASL data were disaggregated by number of years students had spent in their school districts over a period of 3 consecutive school years (1999-2000, 2000-2001, and 2001-2002) and analyzed. Further analyses examined performance of former ELL students now reclassified as fluent English proficient (FEP). Upon completion of analyses, data reports were next prepared and delivered to key administrators in participating school districts. The qualitative component of the study examined what these data suggested to the administrators. To do so, administrators were surveyed by questionnaire and by interview.

Research questions

This study disaggregated WASL performance for ELL students by the years they had in district (from 1 to 5 years) and for former ELL students who had met exit criteria and were now designated as FEP. To examine the data, this study asked the following questions:

1. How does WASL test performance for ELLs differ at 1, 2, 3, 4 and 5 years in district?
2. At what point do ELL scores become similar to the scores of their native English-speaking peers?
3. How do test data for students who have exited the bilingual program compare to those for students still in the program and with other mainstream students?
4. If the districts' ELL scores, grouped together and unreflective of different lengths of time in district, are reported separately from the mainstream, how do the mainstream scores change?

For the second qualitative component, the study sought to discover what administrators thought about the data. Specifically, the questions for this component were:

5. What will administrators from the five districts think about the data?
6. After reviewing the data, what do administrators from the five districts recommend as an equitable way to include ELLs in the state assessment system?

After reviewing the data, what do administrators from the five districts recommend as an equitable way to include ELLs in the state assessment system?

In order to answer questions 5 and 6, administrators were asked these questions: a) What do you think about the data? b) After reviewing the data, what do you recommend as an

equitable way of including ELLs in the state assessment system? c) What do you consider to be your biggest challenge regarding English language learners and WASL achievement?

Research strategy

Sample and settings

Setting. English Language Learner demographics in Washington State vary in many ways. ELL enrollment among the districts ranges from 74% to 0.1% (OSPI, 2002c). Some districts have as many as 70 primary languages while others have as few as one; while about 62% of Washington ELL students speak Spanish as a primary language, approximately 13% speak Russian and Ukrainian. Another 14% speak Vietnamese, Korean, Cambodian, Somali, or Tagalog (OSPI, 2002c). Most districts have an ESL program model, but many with high enrollments of one primary language offer a bilingual program (OSPI, 2002c).

Five southeastern Washington districts of differing characteristics, referred to as Districts A, B, C, D and E, agreed to participate in this study. The districts were chosen for this study because they represented to some extent the demographic variety found in Washington (see Table 4), and combined, they provided a larger database than they would separately. The choice was also due to accessibility and their willingness to participate. While WASL achievement varies in each district (as shown in Table 5), there was no intent to compare districts.

Moreover, programs for ELLs were changing in each of the five districts during the 3 years of this study. Therefore, it would have been difficult to name or compare

program models utilized within the districts. More detailed descriptions of the communities and districts follow.

Table 4
District A-E English Language Learner Primary Languages for 2000-2001

District	Number of students	Percentage of ELLs	Primary languages of students
A	8,970	2.9%	43% Ukrainian and Russian, 20% Spanish, 37% mix of 18 other languages
B	1,600	57%	Spanish
C	821	29%	Spanish
D	13,210	9.5%	84% Spanish, 6% Bosnian, 10% mix of 18 other languages
E	8,185	40.6%	97% Spanish, 3% mix of 11 other languages

Note. Data from Office of Superintendent of Public Instruction (2002c). *Educating limited-English-proficient students in Washington State*. Olympia WA: Office of Superintendent of Public Instruction.

Table 5
Percentages of Fourth Graders in Districts A-E Meeting WASL Standards for 2000-2002

District	Test	Year 1999-2000		Year 2000-2001		Year 2001-2002	
		%	n	%	n	%	n
A	Math	45.7	702	50.9	704	52.4	711
	Reading	70.8	700	72.7	702	66.0	711
	Writing	47.5	697	49.4	694	50.8	702
	Listening	72.9	703	79.8	704	66.8	715
B	Math	14.6	101	9.3	128	17.1	122
	Reading	28.2	99	17.5	124	23.6	121
	Wrlting	15.5	91	15.5	121	15.4	109
	Listening	37.9	100	35.3	135	30.1	123
C	Math	12.3	64	13.4	81	33.8	75
	Reading	27.7	64	36.6	80	37.3	75
	Writing	12.3	64	32.9	80	44.0	74
	Listening	27.7	64	47.6	82	42.7	75
D	Math	44.4	1036	47.1	1008	61.6	1071
	Reading	68.6	1032	67.9	1006	71.9	1065
	Writing	42.5	1031	49.6	994	56.0	1058
	Listening	65.5	1037	74.9	1017	66.3	1077

Table 5 (continued)

E	Math	25.0	707	26.0	702	35.0	730
	Reading	39.9	706	40.0	700	45.1	729
	Writing	22.9	681	24.7	660	34.2	711
	Listening	44.7	715	50.4	705	48.2	741

Note. From Office of Superintendent of Public Instruction, 2002a [On-line], WASL test scores. Available from http://www.k12.wa.us/edprofile/

District A. This community established itself as a center for scientific research following World War II, and since then it has continued to grow. Two university branch campuses, several research labs and one of the nation's major nuclear cleanup sites are located here. While the town's economic base is changing (there are several low-income government housing developments now, and many of the old neighborhoods have become rental areas), it still has the reputation of a professional community. When WASL scores for local districts are posted, this district usually has higher scores. While nearby districts have mainly Spanish-speaking ELLs, District A has a mix of different primary languages groups. Interestingly, this includes children of professional workers (engineers and scientists) on year-long exchange programs as well as refugee families from many parts. With a small percentage of ELLs, District A has used an ESL pullout program, but is now moving toward a content ESL model.

District B. In the heart of orchard country, this town has experienced tremendous growth within the last ten years. The school population has doubled. An influx of Spanish-speaking agricultural workers has brought a 1,600% increase in the district's English language learner population. District B has one of the state's highest poverty levels; in October 2001, 78.3% qualified for free and reduced lunches (OSPI, 2002b). WASL results appear dismal on the surface, among the lowest in the state. However, an outside evaluator hired by the district found that students are making gains in spite of its unique challenges: only 18 to 20% of District B students stay continuously enrolled: no other district in the state has a comparable mobility rate. District B has moved steadily toward a bilingual program with extended day and extended school-year programs, and they are very proud of their long-term students' longitudinal success. (A cohort of 69 students followed from fourth through tenth grade shows encouraging WASL achievement gains; a copy of a bar graph showing these gains is located in Appendix G.)

District C. Like District B, District C is a rural agricultural community where the English language learner population speaks Spanish. Unlike District B, however, this district has less mobility and fewer new students. Many of District C students come from second or third generation immigrant families. In 2000, District C was a recipient of a large 5-year grant from the Bill and Melinda Gates Foundation (n.d.). Receiving the Gates Foundation grant opened the doors for many new partnerships and further grants, for example, with Pacific Science Center, National Science Foundation, and Josten Compass Learning. The Migrant program and University of Washington brought in Twenty First Century Gear Up for sixth through eighth grades. Twenty First Century also provided a new district community learning center; the objective was to involve 75% of

the students, but involvement has been closer to 95%. The center is used mainly for tutorials and homework, computer skills and training, health screening and dental care, family involvement and intervention. Washington Achievers, a high school to college awareness grant earned by 16 state high schools, awarded 30 District C students with full ride college scholarships in 2001-2002. For preschoolers, Even Start and Kinder Camp help with school readiness.

In 2002, all four WASL tests for District C fourth graders showed improvement. At the beginning of this study, District C had an E.S.L. program at the elementary level; it has since moved toward a more bilingual model.

District D. District D is the largest district in this study. This town has a diverse and prospering economy of agriculture, industry, and tourism. It has grown in the last twenty years to about 70,000 people and is still expanding. The district was another recipient of a 5-year Bill and Melinda Gates Foundation grant. It set a goal to get 90% of their students reading on grade level by third grade. In 1996, a district board of the superintendent, a school board member and concerned citizens organized a community foundation to encourage parents to read with their children beginning in infancy for 20 minutes a day. This foundation has created a kindergarten readiness program, free to District D families but copyrighted and marketed to other districts. (Other nearby districts have utilized it.) Scores for fourth graders on the WASL have shown steady improvement in math, reading and writing. Like District A, District D's WASL scores are higher than other area schools', the district has a smaller percentage of ELLs, and while most ELLs are Spanish, there are still sizable numbers of other language groups. District D has also picked up more components of a bilingual program. For instance, Spanish-speaking

students in the primary grades in targeted schools now learn to read in Spanish and are transitioned into English reading in third grade.

District E. District E is an agricultural, manufacturing and transportation center for southeastern Washington. A large number of migrant workers arrive every spring and remain through the summer for a succession of crops; yet many go on to settle here, as in District C. The district has the second-highest Spanish-speaking student count in Washington; it is known for its bilingual education model. A secondary credit exchange program has run for several years, helping students going back and forth yearly between Texas and District E to continue their high school education. Migrant Summer School runs every summer. While the majority of the ELLs are Spanish-speaking, the district also runs a Russian bilingual program. A downtown administration building contains a well-utilized parent outreach center, teaching English as a Second Language, parenting skills and computer skills while offering child care. District E is planning for two dual language programs; one will open next year and another will open the following year.

Sample. While fourth, seventh and tenth grade students are all tested by WASL, fourth grade was the most suited for this study for three reasons. First, school background does not present as much of a variable as it would at seventh or tenth grade. For example, Thomas and Collier (1997) and Ramirez et al. (1991) found that school program model did not make a significant difference at third grade. Yet relationships could be demonstrated between program models and achievement at the seventh grade level (Thomas & Collier, 1997). Second, it was sometimes difficult to ascertain how long a student had been in a district. They may have left and come back, for instance. Or records may have been inaccurate. In middle or high school, there would have been even

more years of school records to investigate. Third, by looking at fourth grade, we looked at the group most representative of ELLs in Washington and in the U.S. Nearly half of Washington's ELLs are in grades K-3 (OSPI, 2003a); 67% of the nation's school-aged ELLs are at the elementary school level (Kindler, 2002).

By collecting data from the five districts, this study could both represent a geographic area while enlarging the sample. (Table 6 shows sample sizes.) Further, three of the districts have ELL subgroup sizes less than 100, reducing their individual statistical reliability (Hill & De Pascale, 2003).

Table 6.
Sample sizes for Districts A-E, 2000-2002

District	1999-2000		2000-2001		2001-2002		All 3 years	
	ELLs	All 4th	ELLs	All 4th	ELLs	All 4th	ELLs	All 4th
A	14	703	17	704	15	715	46	2,122
B	55	101	87	135	70	123	212	359
C	28	84	36	83	25	75	89	242
D	90	1,037	115	1,017	116	1,077	321	3,131
E	331	715	380	705	362	741	1,073	2,161
Combined	518	2,640	635	2,644	588	2,731	1,741	8,015

Note. N's for all fourth grade (listening) from Office of Superintendent of Public Instruction, 2002a [On-line], WASL test scores. Available from http://www.k12.wa.us/edprofile/. ELL n's provided by Districts A-E.

Strategies

Quantitative component. The quantitative component required further statistical analyses of fourth grade ELL WASL data. The five school districts were given a 6-page instrument on which to record fourth grade WASL passing percentages for ELL students by subgroup. (This instrument is found in Appendix A; for completed instruments from the five districts, see Appendix B). There were six subgroups. Five subgroups represented different lengths of stay in district (1 year, 2 years, 3 years, 4 years, and 5 years) and a sixth group of students represented those previously enrolled in the bilingual program but now exited from the program. (Students exited the bilingual program upon reaching the 35th percentile or above on reading and language portions of a norm-referenced achievement test, such as the California Test of Basic Skills or the Iowa Test of Basic Skills. Since the time of this study, Washington has changed to the Washington Language Proficiency Test for its yearly language proficiency assessment required under NCLBA guidelines).

This study also investigated differences with ELL scores included and ELL scores removed, as a source of possible further information. For example, it could be that percentages meeting standard could be very similar with the ELL scores removed. On the other hand, it might be that scores could change considerably.

Qualitative component. Following the analyses, a report of the findings (see Appendix C) was sent to the administrators in each district who were responsible for the fourth grade language assistance program. In most instances, these were the superintendent, elementary education administrator and bilingual program director. However, there were exceptions. For example, smaller districts such as District B and C

did not have these three positions. In District E, a new superintendent was not aware of the study and was extremely busy, so the assistant superintendent participated. In any case, these administrators were the district figures most likely to utilize these assessment data. (Table 7 lists the participating administrators.) A five-item questionnaire accompanied the report. This report was sent within 2 weeks following completion of data analysis. Within approximately 2 more weeks, a follow-up discussion was scheduled in each district where possible with the key administrators. (This happened in three

Table 7
District administrators participating in interviews

District	Participant
A	Superintendent
	Director of Elementary Education
	Director of State and Federal Programs
B	Superintendent
	Assistant Superintendent
C	Superintendent
	Literacy and Assessment Director
D	Superintendent
	Director of Elementary Education
	Director of Bilingual Education
E	Assistant Superintendent
	Director of Special Programs

districts. In two districts, discussion had to be held individually and, in the case of one semi-retired administrator, responses were given in writing with some follow up discussion over email.) In the meetings, which typically lasted 30 minutes, the first three questions from the questionnaire were discussed, and in many cases we went on to discuss all five questions. The questions were designed to probe the administrators' thoughts about the findings and to ascertain if their thoughts reflected understanding of the need for time for ELLs to acquire test readiness. Also, administrators were questioned what they perceived as the biggest challenge regarding ELLs and the fourth grade WASL. The questions were:

1. Now that you have studied the data, what do you think about the findings?
2. After seeing the data, what do you think would be the most equitable way to include ELLs in the state testing system?
3. What is your biggest challenge in regards to English language learners and the fourth grade WASL?

Two further questions followed. They could be answered individually at a later time, if more time was needed for reflection, or if the participant wanted to answer privately. These questions could be answered in writing or over the telephone. Most participants chose to answer them at the interview, one answered by email, and one answered by telephone.

4. Is there anything you want to say additionally about these data?
5. Is there anything you want to say additionally about the state testing system?

The responses were recorded, except in one case. In that case, the superintendent did not want to be tape-recorded. He preferred that we just "have a conversation". Later, answers from all interviews were transcribed, studied and examined for themes.

Data collection

Quantitative data collection. Originally, the superintendents in each district were contacted to be part of the study. Each superintendent who was contacted expressed interest in the basis for the study and agreed to take part. The next step was for them to assign someone to perform the data collection.

Collecting the quantitative data for this study proved to be more difficult than expected. One district, District B, was able to get the data immediately. This district, with the state's second lowest WASL scores, had already done intensive data analysis. They had hired an analyst to study their district and achievement data to help explain their low performance, and according to the superintendent, they had analyzed their statistics and data in every way imaginable. It was not so easy for other districts. District C was the only one with computer capability to access the needed information readily, and that may have been because there was only one elementary school. Even so, there was only one person who knew how to do it. She was the literacy and assessment director, hired with money from the Bill and Melinda Gates Foundation Grant. The other districts did not have length of stay data entered into district databases. District E (with the highest number of ELLs) had the hardest time and took the longest to complete data collection. There were many records to go through, and students had moved around between buildings. Many of the students were from migrant families; they had left the district and returned a month or several months later. For Districts A, D and E, gathering

the data was a student-by-student process, and it was often a building-by-building process as well. While Districts B and C provided their data relatively quickly, the other three districts took much longer than expected. In Districts D and E, the bilingual program directors were the ones doing the actual data collecting; they both stated that the records were too jumbled and the process too intricate for anyone else to have to sort through it. For District E, this meant combing through more than 900 student records.

Data analysis

Quantitative data analysis. After collecting data from each district, a larger data-base was built, using Excel. This resulted from combining data from the five different school districts for each of the 3 school years. Passing rate percentages could then be assigned to each length of stay group. Charts combining data for each school year were prepared, and a chart combining data for all 3 school years was also completed.

Qualitative data analysis. After recording and transcribing the questionnaire answers and roundtable discussions with the district administrators, the coding of responses (Glesne, 1999) and data displays (Miles & Huberman, 1994) followed. (Appendix D contains a list of interviews and interviewees; transcribed interviews are in Appendix E.) In this manner, themes emerged about what administrators in school districts thought about these data, what recommendations they had for a more equitable way of including ELLs in the state assessment process, and whether having more time for WASL readiness was part of their thinking (see Appendix F for a list of themes and sub-themes). These analyses will be discussed in more detail in the following chapter.

Chapter 4

Results

The primary purpose of this study was to seek and describe a relationship between fourth grade ELL lengths of stay and test performance on the Washington Assessment of Student Learning. This included the performance of former bilingual program students now reclassified as fluent English proficient. This purpose was addressed with a quantitative component. The secondary purpose, to examine administrator perceptions through administrator feedback to the quantitative component, was addressed through a qualitative component. This chapter will contain descriptions and findings of each component.

Quantitative component

Description

Five neighboring districts in southeastern Washington participated in this study. They contributed testing data from the Washington Assessment of Student Learning for fourth grade ELLs in their districts, for 3 consecutive school years (1999-2000, 2000-2001, and 2001-2002). The districts disaggregated the percentages of ELLs meeting standard by years in district (Table 8 shows test data before disaggregation, along with aggregate ELL data from Washington State). Also, they contributed data for students formerly classified as ELLs but now redesignated as FEP. For each school year, there were subgroups for students with 1 year or less in district, from 1 to 2 years in district, from 2 to 3 years in district, from 3 to 4 years in district and from 4 to 5 years or more in

district. (This last group would be students who entered in kindergarten; some who might have repeated a year.) Test data for all fourth graders were also included.

Table 8
Combined percentages of ELLs meeting standard from Districts A-E and percentage of state ELLs meeting standard, school years 1999-2000, 2000-2001, 2001-2002

	1999-2000		2000-2001		2001-2002	
Test	Districts A-E	State	Districts A-E	State	Districts A-E	State
Math						
%	7.1%	10.9%	8.7%	11.6%	14.8%	18.2%
n=	518	2,215	635	4,151	588	4,491
Reading						
%	18.0%	20.9%	20.6%	24.0%	22.8%	24.8%
n=	518	2,191	635	4,129	588	4,468
Writing						
%	10.8%	10.5%	12.3%	15.4%	14.5%	19.4%
n=	518	2,217	635	4,177	588	4,524
Listening						
%	26.8%	30.0%	34.5%	36.5%	27.7%	35.4%
n=	518	2,113	635	3,969	589	4,319

Note. District A-E data provided by districts A-E. Data from the exited group are not included in this table. State data from Office of Superintendent of Public Instruction, 2002 [On-line], WASL test scores. Available from http://www.k12.wa.us/edprofile/

Once these data were gathered from the five school districts, they were entered onto Excel spreadsheets and combined to make larger samples for each of the school

years. Charts were produced for each school year. Finally, for an even larger sample, all 3 years were combined (see Figures 1-4 for charts).

Results

In this section, general observations will be discussed first. Because each test showed different achievement results, more specific consideration of each of the four tests (Math, Reading, Writing and Listening) will follow.

General observations. Some generalizations could be drawn across all 3 of the testing years. First, the districts' combined total ELL WASL performance (Table 8) was lower than the state's. Also, percentages meeting standard did rise over the three testing years, in both the combined Districts A-E sample and in the state ELL population, other than in Listening in 2001-2002.

In Districts A-E, bilingual program students in their district for under 2 years generally exhibited the lowest performance of the groups. These first 2 years in district often showed similar achievement; the second year was not necessarily higher than the first year. Sometimes, students in the second year performed lower than the first year group. Performance for students in the district for 3 years tended to be higher, especially in Reading and Listening. In the fourth and fifth year groups, achievement was generally higher than the first 3 years, but not always. Again, as in the first and second year in district, the fourth and fifth year group could often be grouped together: at times, the fourth year group outperformed the 5 year group, and the fifth year did not show a steady rise over the fourth year.

1999-2000								
MATH	**<1 year**	**1-2 years**	**2-3 years**	**3-4 years**	**4-5+ yrs**	**exited**	**all 4th**	**non-prog.**
Dist./ ELL #								
A 14	0	1	0	0	0	3	321	320
B 55	0	0	0	0	0	2	15	15
C 28	0	0	1	0	4	n/a	11	6
D 90	2	1	1	0	2	0	460	454
E 331	0	2	1	2	20	19	177	152
ttl	**2**	**4**	**3**	**2**	**26**	**24**	**984**	**947**
518 n	38	55	39	42	344	97	2,629	2,111
	5%	2%	8%	2%	8%	25%	37%	45%
READING	**<1 year**	**1-2 years**	**2-3 years**	**3-4 years**	**4-5+ yrs**	**exited**	**all 4th**	**non-prog.**
Dist./ ELL #								
A 14	0	2	0	0	0	6	495	493
B 55	0	0	0	1	3	6	29	25
C 28	0	0	1	0	7	n/a	31	23
D 90	2	1	0	2	8	6	708	695
E 331	0	0	2	7	57	40	282	216
ttl	**2**	**3**	**3**	**10**	**75**	**58**	**1,545**	**1,452**
518 n	38	55	39	42	344	97	2,623	2,105
	5%	5%	8%	24%	22%	60%	59%	69%
WRITING	**<1 year**	**1-2 years**	**2-3 years**	**3-4 years**	**4-5+ yrs**	**exited**	**all 4th**	**non-prog.**
Dist./ ELL #								
A 14	0	1	0	0	0	5	331	330
B 55	0	0	0	0	0	4	16	16
C 28	0	0	1	0	8	n/a	28	19
D 90	1	1	1	1	4	4	438	430
E 331	1	2	1	2	32	17	156	118
ttl	**2**	**4**	**3**	**3**	**44**	**30**	**969**	**913**
518 n	38	55	39	42	344	97	2,594	2,076
	5%	7%	8%	7%	13%	31%	37%	44%
LISTENING	**<1 year**	**1-2 years**	**2-3 years**	**3-4 years**	**4-5+ yrs**	**exited**	**all 4th**	**non-prog.**
Dist./ ELL #								
A 14	0	2	0	0	0	6	512	510
B 55	0	0	0	2	11	5	39	26
C 28	1	0	1	0	7	n/a	40	31
D 90	5	3	3	2	11	7	679	655
E 331	2	5	3	9	72	35	320	229
ttl	**8**	**10**	**7**	**13**	**101**	**53**	**1,590**	**1,451**
518 n	38	55	39	42	344	97	2,640	2,122
	21%	18%	18%	31%	29%	55%	60%	68%

Figure 1
Percent of ELLs/ exited students meeting standards in Districts A-E, 1999-2000

2000-2001								
MATH	**<1 year**	**1-2 years**	**2-3 years**	**3-4 years**	**4-5+ yrs**	**exited**	**all 4th**	**non-prog.**
Dist./ ELL #								
A 17	1	0	0	0	0	3	359	358
B 87	0	0	0	0	0	7	12	12
C 36	0	0	1	2	3	n/a	11	5
D 115	0	0	3	0	4	2	475	468
E 380	0	5	1	7	28	21	183	142
ttl	**1**	**5**	**5**	**9**	**35**	**33**	**1,040**	**985**
635 n	55	63	68	78	371	113	2,632	1,997
	2%	8%	7%	12%	9%	29%	40%	49%
READING	**<1 year**	**1-2 years**	**2-3 years**	**3-4 years**	**4-5+ yrs**	**exited**	**all 4th**	**non-prog.**
Dist./ ELL #								
A 17	2	0	0	0	1	4	510	507
B 87	0	0	0	0	1	12	21	20
C 36	1	0	1	3	6	n/a	31	20
D 115	3	0	4	6	14	6	683	656
E 380	0	5	6	12	66	42	280	191
ttl	**6**	**5**	**11**	**21**	**88**	**64**	**1,525**	**1,394**
635 n	55	63	68	78	371	113	2,628	1,993
	11%	8%	16%	27%	24%	57%	58%	70%
WRITING	**<1 year**	**1-2 years**	**2-3 years**	**3-4 years**	**4-5+ yrs**	**exited**	**all 4th**	**non-prog.**
Dist./ ELL #								
A 17	1	0	0	0	1	4	343	341
B 87	0	0	0	0	3	12	23	20
C 36	0	0	0	4	6	n/a	28	18
D 115	2	1	4	3	3	2	493	480
E 380	0	0	5	4	41	23	163	113
ttl	**3**	**1**	**9**	**11**	**54**	**41**	**1,050**	**972**
635 n	55	63	68	78	371	113	2,566	1,931
	5%	2%	13%	14%	15%	36%	41%	50%
LISTENING	**<1 year**	**1-2 years**	**2-3 years**	**3-4 years**	**4-5+ yrs**	**exited**	**all 4th**	**non-prog.**
Dist./ ELL #								
A 17	1	1	1	0	1	3	562	558
B 87	0	1	1	4	14	11	43	23
C 36	1	1	2	8	6	n/a	40	22
D 115	4	3	7	7	25	7	762	716
E 380	3	8	8	16	96	46	355	224
ttl	**9**	**14**	**19**	**35**	**142**	**67**	**1,762**	**1,543**
635 n	55	63	68	78	371	113	2,644	2,009
	16%	22%	28%	45%	38%	59%	67%	78%

Figure 2
Percent of ELLs/ exited students meeting standards in Districts A-E, 2000-2001

2001-2002								
MATH	**<1 year**	**1-2 years**	**2-3 years**	**3-4 years**	**4-5+ yrs**	**exited**	**all 4th**	**non-prog.**
Dist./ ELL #								
A 15	0	0	2	0	0	4	372	370
B 70	0	0	1	0	2	9	20	17
C 25	0	0	0	0	1	n/a	26	25
D 116	1	2	1	6	13	4	660	637
E 362	1	3	6	12	36	35	256	198
ttl	**2**	**5**	**10**	**18**	**52**	**52**	**1,334**	**1,247**
588 n	47	58	71	84	328	116	2,710	2,122
	4%	9%	14%	21%	16%	45%	49%	59%
READING	**<1 year**	**1-2 years**	**2-3 years**	**3-4 years**	**4-5+ yrs**	**exited**	**all 4th**	**non-prog.**
Dist./ ELL #								
A 15	0	1	3	0	0	5	469	465
B 70	0	0	0	0	5	14	29	24
C 25	0	0	0	0	3	n/a	28	25
D 116	1	2	3	6	20	5	765	730
E 362	1	4	8	18	56	52	328	241
ttl	**3**	**9**	**14**	**24**	**84**	**76**	**1,619**	**1,485**
588 n	47	58	71	84	328	116	2,702	2,114
	6%	16%	20%	29%	26%	66%	60%	70%
WRITING	**<1 year**	**1-2 years**	**2-3 years**	**3-4 years**	**4-5+ yrs**	**exited**	**all 4th**	**non-prog.**
Dist./ ELL #								
A 15	0	0	2	0	0	5	356	354
B 70	0	0	0	0	1	10	19	18
C 25	0	0	0	0	5	n/a	33	28
D 116	2	3	0	5	12	4	592	570
E 362	2	2	3	12	36	39	243	188
ttl	**4**	**5**	**5**	**17**	**54**	**58**	**1,243**	**1,158**
588 n	47	58	71	84	328	116	2,667	2,079
	9%	9%	7%	20%	16%	50%	47%	56%
LISTENING	**<1 year**	**1-2 years**	**2-3 years**	**3-4 years**	**4-5+ yrs**	**exited**	**all 4th**	**non-prog.**
Dist./ ELL #								
A 15	1	1	5	0	2	2	478	469
B 70	0	0	1	1	11	13	37	24
C 25	0	0	0	0	4	n/a	32	28
D 116	1	2	2	9	20	5	714	680
E 362	2	4	13	13	71	49	357	254
ttl	**4**	**7**	**21**	**23**	**108**	**69**	**1,618**	**1,455**
588 n	47	58	71	84	328	116	2,731	2,143
	9%	12%	30%	27%	33%	59%	59%	68%

Figure 3
Percent of ELLs/ exited students meeting standards in Districts A-E, 2001-2002

All 3 years combined								
MATH	**<1 year**	**1-2 years**	**2-3 years**	**3-4 years**	**4-5+ yrs**	**exited**	**all 4th**	**non-prog.**
District								
A 46	1	1	2	0	0	10	1052	1048
B 212	0	0	1	0	2	18	47	44
C 89	0	0	2	2	8	n/a	48	36
D 321	3	3	5	6	19	6	1,595	1,559
E 1,073	1	10	8	21	84	75	616	492
ttl	**5**	**14**	**18**	**29**	**113**	**109**	**3,358**	**3,179**
1,741n	140	176	178	204	1,043	326	7,971	6,230
	4%	8%	10%	14%	11%	33%	42%	51%
READING	**<1 year**	**1-2 years**	**2-3 years**	**3-4 years**	**4-5+ yrs**	**exited**	**all 4th**	**non-prog.**
District								
A 46	2	3	3	0	1	15	1,474	1,465
B 212	0	0	0	1	9	32	79	69
C 89	1	0	2	3	16	n/a	90	68
D 321	6	3	7	14	42	17	2,156	2,081
E 1,073	1	9	16	37	179	134	890	648
ttl	**10**	**15**	**28**	**55**	**247**	**198**	**4,689**	**4,331**
1,741n	140	176	178	204	1,043	326	7,953	6,212
	7%	9%	16%	27%	24%	61%	59%	70%
WRITING	**<1 year**	**1-2 years**	**2-3 years**	**3-4 years**	**4-5+ yrs**	**exited**	**all 4th**	**non-prog.**
District								
A 46	1	1	2	0	1	14	1,030	1,025
B 212	0	0	0	0	4	26	58	54
C 89	0	0	1	4	19	n/a	89	65
D 321	5	5	5	9	19	10	1,523	1,480
E 1,073	3	4	9	18	109	79	562	419
ttl	**9**	**10**	**17**	**31**	**152**	**129**	**3,262**	**3,043**
1,741n	140	176	178	204	1,043	326	7,827	6,086
	6%	6%	10%	15%	15%	40%	42%	50%
LISTENING	**<1 year**	**1-2 years**	**2-3 years**	**3-4 years**	**4-5+ yrs**	**exited**	**all 4th**	**non-prog.**
District								
A 46	2	4	6	0	3	11	1,552	1,537
B 212	0	1	2	7	36	29	119	73
C 89	2	1	3	8	17	n/a	112	81
D 321	10	8	12	18	56	19	2,155	2,051
E 1,073	7	17	24	38	239	130	1,032	707
ttl	**21**	**31**	**47**	**71**	**351**	**189**	**4,970**	**4,449**
1,741n	140	176	178	204	1,043	326	8,015	6,274
	15%	18%	26%	35%	34%	58%	62%	71%

Figure 4
Percent of ELLs meeting standards in Districts A-E, all 3 school years combined (1999-2000, 2000-2001, 2001-2002)

Testing achievement patterns in the study seemed to follow De Avila's 1997 findings on the progression of ELL language development. He found that ELLs entering school with few English skills progressed during the first few years in English listening and speaking. Reading followed, beginning in the third year of elementary school. Writing lagged behind, beginning to build toward the end of elementary school (De Avila, 1997). Similarly, performance on the WASL Listening test was higher for ELLs in their first 2 years, but in Reading, achievement did not begin to rise until the third year. Reading scores were considerably higher than Math and Writing. Writing is utilized to a great extent in the WASL mathematics responses; these two areas were the least successful. Yet they did show a slow improvement, more noticeable at the fourth and fifth year in district.

Students who exited the bilingual program showed a bigger leap forward in achievement gains than any yearly subgroup compared to another. Consistently, FEPs gained close to 20% or more than the 5-year-in-district students, in all four areas. The highest gains were in Reading, where FEP achievement grew by 38% in 1999-2000, 33% in 2000-2001, 40% in 2001-2002 and 37% in the combined sample. Fluent English proficient students performed close to their district's fourth grade group overall, the furthest point being 8 percentage points behind, except in the case of Mathematics in 1999-2000 (12 percentage points) and 2000-2001 (11 percentage points). In some cases, the percentage of exited students meeting standard was higher than the percentage of all the fourth grade students (Writing, 2001-2002, and Reading, 1999-2000 and 2001-2002 and combined Reading scores for all 3 years). For Listening 2001-2002, performance for exited students and all students was equal.

The study also looked to see how district scores would change with the ELL student scores removed. Percentages meeting standard did rise in every case. In Math, the percentage meeting standard rose by almost 10% in each of the testing years. In Reading, the increase in percentage meeting standard varied from 10% to 12% without the ELL scores included. The percentage meeting standard in Writing varied from 7% to 9% when bilingual program scores were removed. In Listening, the increase in percentage meeting standard when ELLs were removed also rose from 7% to 9% over the 3 years.

Math. In the area of Math, students showed improvement, although in an inconsistent manner (see Figure 5). For example, in 1999-2000, the percentage passing in the second year was lower than in the first. The fourth year performance was lower than the third. In 2000-2001, the percentage passing in the third year was lower than the second. Students in the bilingual program for 4 years scored higher than those in the program for 5 years in years 2000-2001 and 2001-2002. At the same time, scores did rise from one year to the next. All fourth grade and non-program fourth grade scores rose too. (Still, Math continues to be the WASL test with the lowest passing percentages, for ELLs and non-ELLs alike.) Exited students' performance stayed further behind all fourth grade students than it did for the other three tests (Reading, Writing and Listening), although the gap was only 4% in 2001-2002, as seen in Figures 1 through 3. Taking the bilingual program scores out of District A-E total scores raised the percent meeting standard by 12% in 1999-2000, 9% in 2000-2001 and 10 % in 2001-2002.

Math performance

Time in district	1999-2000	2000-2001	2001-2002	All 3 years
<1 year	5%	2%	4%	4%
1-2 years	2%	8%	9%	8%
2-3 years	8%	7%	14%	10%
3-4 years	2%	12%	21%	14%
4-5+ years	8%	9%	16%	11%
Exited from program	25%	29%	45%	33%
All fourth	37%	40%	49%	42%
Non-program fourth	45%	49%	59%	51%

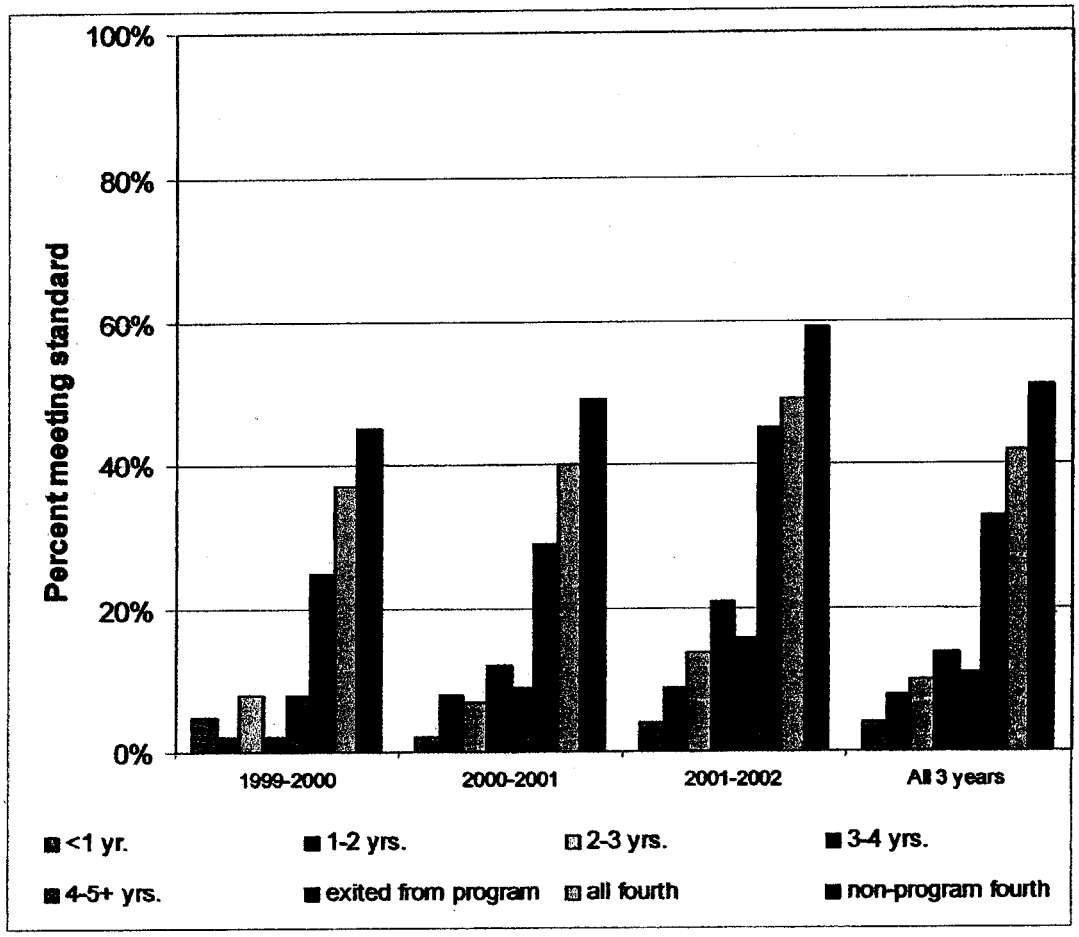


Figure 5
Percent of fourth grade ELLs meeting standard in Math on Washington Assessment of Student Learning, Districts A-E, across 3 school years

Reading. In Reading, there was improvement each testing year for students in the bilingual program in each of the subgroups, except for one: the first year in district group between 2000-2001 and 2001-2002, as shown in Figure 6. (Students in the program for 2, 3, 4, and 5 years showed better performance each year than they had the year before; i.e. students in the 3-year group scored better in 2002 than they did in 2001, and they scored better in 2001 than they did in 2000.) Interestingly, students in the bilingual program for 4 years scored higher than students in the program for 5 years, for all 3 testing years. In Reading, there was a consistent jump of about 40% in performance between the bilingual program student scores and the exited/ all fourth/ non-program fourth grade student scores (see Figure 6). This was the biggest jump of all four tests between students in program and exited students. Further, exited students performed at higher or almost same percent as all fourth graders, for all 3 years. Finally, taking the bilingual program scores out raised the percent meeting standard by 10% for 1999-2000, 12% for 2000-2001, and 10% for 2001-2002.

Curiously, all fourth grade students and all non-program fourth graders in the study seemed to have hit a plateau. Likewise, all fourth graders in Washington showed a group flat rate of growth for the 3 testing years (see Table 3). In testing year 1999-2000, 65.8% of fourth graders met standard in reading. In 2000-2001, 66.1% met standard; in 2001-2002, 65.6% did so. The trend appears to continue: in testing year 2002-2003, OSPI (2003c) reported that 66.7% of all fourth graders met standard.

Reading performance

Time in district	1999-2000	2000-2001	2001-2002	All 3 years
<1 year	5%	11%	6%	7%
1-2 years	5%	8%	16%	9%
2-3 years	8%	16%	20%	16%
3-4 years	24%	27%	29%	27%
4-5+ years	22%	24%	26%	24%
Exited from program	60%	57%	66%	61%
All fourth	59%	58%	60%	59%
Non-program fourth	69%	70%	70%	70%

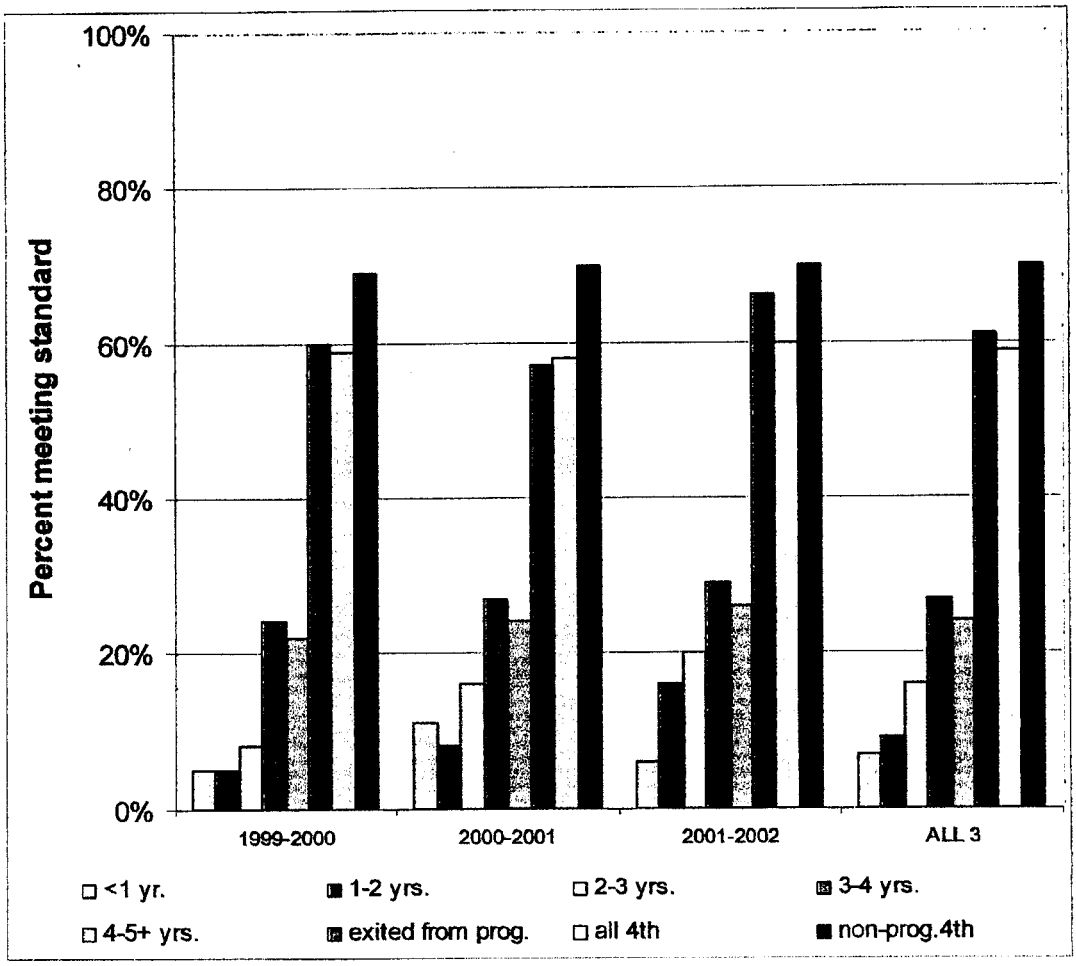


Figure 6
Percent of fourth grade ELLs meeting standard in Reading on Washington Assessment of Student Learning, Districts A-E, across 3 school years

Writing. Improvement over the 3 years was not consistent for ELLs in district for 1, 2 and 3 years (Figure 7). For bilingual program students in district 4 and 5 years, there was an increase in performance. Yet performance for bilingual program students stayed at 20% or below. Exited students showed a more dramatic growth in achievement, from 31% to 36% to 50% over the 3 years. All fourth grade and non-program fourth grade scores rose too. The percentage of exited students who met standard surpassed the percentage of all fourth graders meeting standard in Writing in 2001-2002. Moreover, exited students' performance was only 5 percentage points behind all fourth graders for 1999-2000 and 2000-2001. With Math, achievement in Writing for all groups of students still in the bilingual program was considerably lower than reading and listening. Without the bilingual program scores, the percent meeting standard would rise by 7% in 1999-2000. It would rise by 9% in 2000-2001 and by 9% in 2001-2002.

Listening. The highest performance was found here, for both students in the bilingual program and the non-program students. However, it was highest in 2000-2001 and decreased in 2001-2002 (Figure 8). The percent of students in the bilingual program meeting the listening standard was often 10 to 20 points beyond the same group in other WASL tests in the same year. For instance, in 1999-2000, students in district for 1 year met standard at a rate of 21% in listening, but 5% in math, 5% in reading, and 5% in writing. Students in district for 3-4 years, in the same year, met standard in listening at 31%, with 2% in math, 24% in reading, 7% in writing. (Again, in De Avila's 1997 study, ELLs' listening comprehension built first, rapidly for two years, followed by reading, with writing even later.) For the 1999-2000 and 2000-2001 Listening test, students in their district 4 years outperformed those in their district for 5 years. Exited students

Writing performance

Time in district	1999-2000	2000-2001	2001-2002	All 3 years
<1 year	5%	5%	9%	6%
1-2 years	7%	2%	9%	6%
2-3 years	8%	13%	7%	10%
3-4 years	7%	14%	20%	15%
4-5+ years	13%	15%	16%	15%
Exited from program	31%	36%	50%	40%
All fourth	37%	41%	47%	42%
Non-program fourth	44%	50%	56%	50%

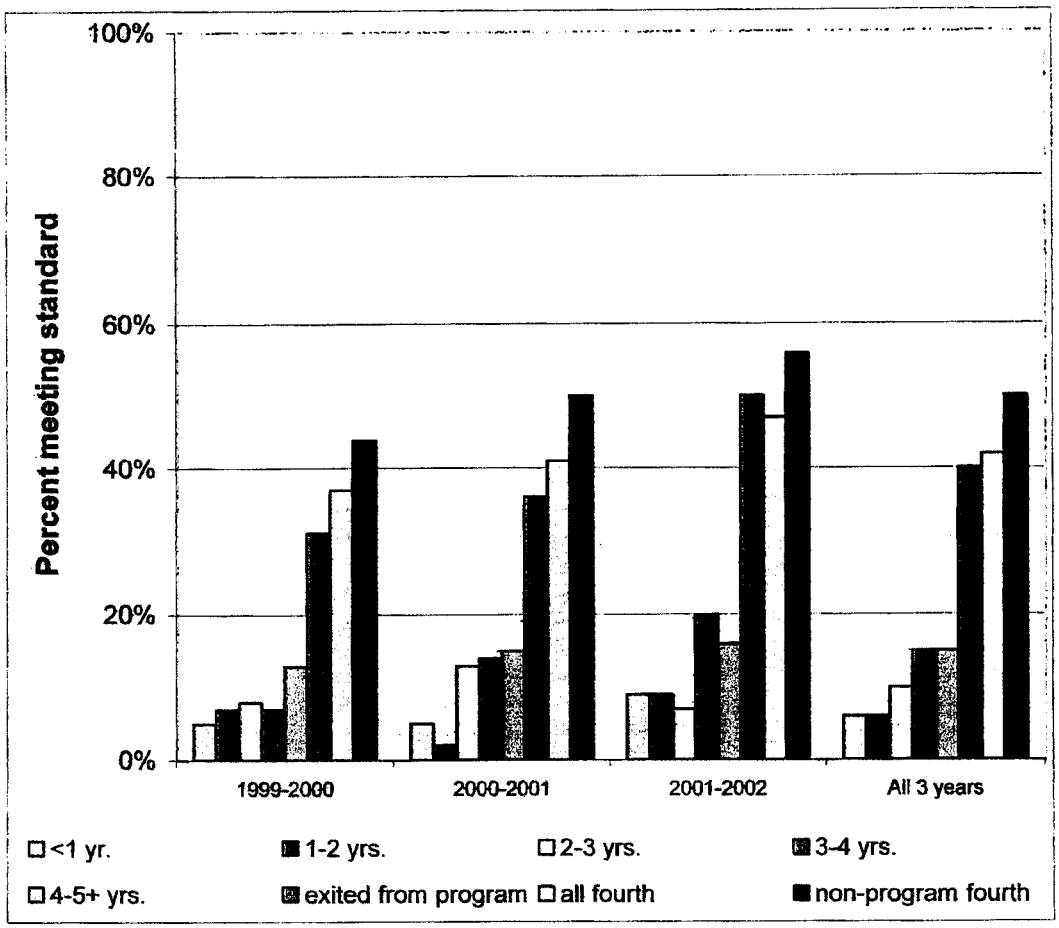


Figure 7
Percent of fourth grade ELLs meeting standard in Writing on Washington Assessment of Student Learning, Districts A-E, across 3 school years

Listening performance

Time in district	1999-2000	2000-2001	2001-2002	All 3 years
<1 year	21%	16%	9%	15%
1-2 years	18%	22%	12%	18%
2-3 years	18%	28%	30%	26%
3-4 years	31%	45%	27%	35%
4-5+ years	29%	38%	33%	34%
Exited from program	55%	59%	59%	58%
All fourth	60%	67%	59%	62%
Non-program fourth	68%	78%	66%	71%

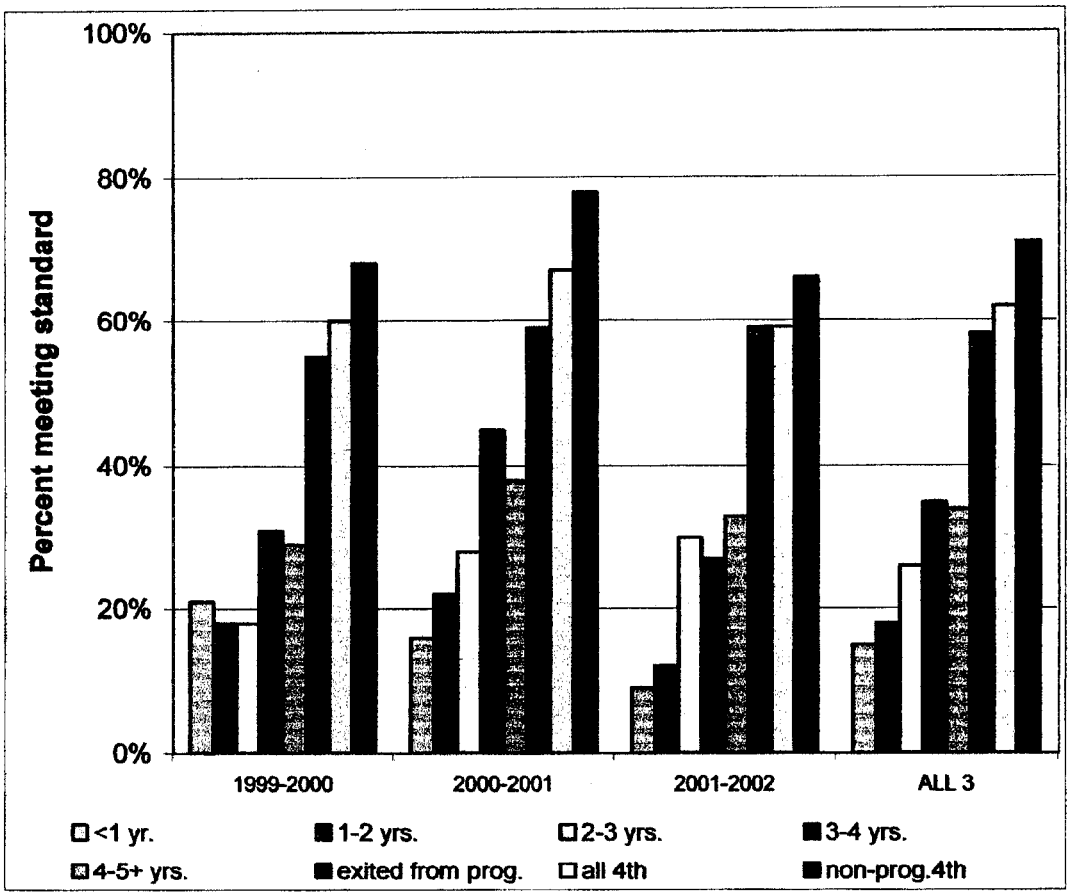


Figure 8
Percent of fourth grade ELLs meeting standard in Listening on Washington Assessment of Student Learning, Districts A-E, across 3 school years

performed equally with all students in 2001-2002. (In 2001-2002, listening scores went down all across the state.) Taking out the bilingual program scores would raise the percent meeting standard by 8% in 1999-2000, 9% in 2000-2001, and 7% in 2001-2002.

Quantitative component summary. Disaggregating WASL data for fourth grade English language learners by number of years they had spent in district provided a picture of slow progress. In this study, it was possible to divide students into five lengths-of-stay groups. Groups did not show consistent year-by-year improvement; in fact, there were plateaus and areas of backsliding. Generally, first and second year groups did not perform differently. The third year group seemed to pull ahead. Fourth and fifth year groups, like the first and second year groups, did not perform differently on the whole. While there was a general trend upwards across the five length of stay groups, in some cases it was very slight. For instance, in Mathematics for school year 1999-2000, ELLs with 1 year in district met standard at a rate of 5%; ELLs with 5 years in district met standard at 8%. In Writing for 2001-2002, first year students met standard at 9% while the fifth year group met standard at 16%. Overall, students were the most successful on the Listening test, followed by Reading. ELLs had less success on Writing and Math. This progression echoes the English language development identified by De Avila (1997).

Students in this study who exited from the bilingual program were able to achieve at a much higher level. In many instances, they were close or at the level of their native English-speaking peers, although in the area of Math there was still a gap of 15 to 20%. However, all Washington fourth graders score continue to score lowest on the Math portion of WASL (OSPI, 2002a). In Reading, the level of achievement for exited students was at or above that of all fourth graders in districts A-E. Finally, removing the

bilingual program scores from the district scores usually resulted in a difference of about 10% more meeting standard.

Qualitative component

Description. Following data analysis, data summary reports were prepared for each participating administrator in the five districts (see Appendix C). The reports began with a description of the study along with the study's purpose and study questions. Data were presented in the form of bar graphs and charts. Findings were summarized, and a copy of the open-ended questions for the follow-up interview was included. Interview questions were developed along guidelines suggested by Glesne (1999).

Each administrator was contacted for an appointment for the follow-up interview. In three districts, an interview was scheduled with all participating administrators present. In the other two, it was not possible, and interviews were conducted separately. One interview had to be conducted by email (the administrator was in the process of retiring and extremely busy helping hire and train a replacement). All interviews took place within 2 to 4 weeks after administrators received their data summary reports (a schedule of interviews and interviewees is in Appendix D). Interviews were tape recorded, with the exception of District B. This particular superintendent did not wish to be tape-recorded, saying he "just preferred to have a conversation." (As it happened, he had not had time to read the data summary report yet and needed to go over it first.)

Interviews lasted from 20 minutes (District B, where a crisis of some sort was occurring) to 45 minutes (with the elementary education director in District D and the administrators in District E). After interviewing administrators, transcripts were prepared (found in Appendix E). Transcripts were next examined for themes. In an emergent style

suggested by Erickson (1986) and Wolcott (1994), themes were not pre-anticipated but allowed to surface on their own. Participants' responses were coded onto a display (Appendix F) in a procedure recommended by Miles and Huberman (1994). For validation, two colleagues also studied the transcripts and tentative themes and added feedback and suggestions.

Results

Participants. None of the participants were new to their positions. However, their experience with ELL education varied considerably. In District A (with only 3% of the district in the bilingual program), the superintendent and elementary education director had no background in ELL education and very little to do with their district's bilingual program. The state and federal programs director did have some past experience as an ESL teacher, along with experience as a Special Education and Title One teacher, yet the district ESL program was only 10% of her job. On the other hand, in District B where 57% of the students were in the bilingual program, both administrators had many years of experience in ELL education. The superintendent had been there 10 years, and the assistant superintendent had worked both as teacher and principal in bilingual schools for many years. The superintendent and literacy and assessment director in District C (27% bilingual program students) also both had a great deal of professional development and experience in ELL education. (They are both currently working on doctorates, focusing on issues of assessment, administration and bilingual education.) In District D, the superintendent had written his dissertation many years previously on the effects of teaching Navajo children to read first in Navajo. He was very cognizant of the benefits of strengthening L1 to aid acquisition of English. The elementary education director did not

have a background in bilingual education, but he had begun attending inservices and conferences to become better informed. Both the migrant/ bilingual program director of District D and the director of special programs in District E came from Spanish-speaking families. They had both worked first as paraprofessionals, gone to college for teaching certificates, and worked as teachers in their districts before becoming program directors.

Some of the participants, notably administrators from Districts A and E, are currently involved in trying to change some aspect of the state assessment system. The superintendent and colleagues from District A are campaigning to replace the planned addition of further WASL-like assessments in 2005 with an alternative instrument. (Washington is developing tests referred to as "Baby WASLs" for third, fifth, sixth, and eighth grades, plus one high school grade, to meet the requirement of the *No Child Left Behind Act* for further tests at these grades.) District A's superintendent strongly believes that the Measure of Academic Progress (MAP) test (developed by Northwest Evaluation Association, n.d.) is more suitable, more time and cost effective, and more useful. He argues that this computerized test, taken individually, takes only 30 to 60 minutes and is more cost effective. It can immediately inform the teacher about the learning level of each student in math, reading and language, thus potentially informing instruction and allowing appropriate differentiation for ELLs and other diverse learners. Several states use MAP as part of their assessment system. With the backing of the school board, the superintendent has put together a coalition of supporters to lobby for this change. District C's superintendent and literacy and assessment director are among other supporters in this effort.

In District E, the assistant superintendent and the special programs director are on the state AYP and Bilingual/ Special Education Committee. This committee is building a case to keep ELLs and Special Education students from automatic participation in state assessments. The assistant superintendent believes that by definition of being in programs identified for special assistance, students in Special Education and the state bilingual program should not be tested with the mainstream. "How do you test them equitably with everybody else? You can't." The special programs director is on the English Language Development Standards Committee and the Bilingual State Assessment Committee as well. The Assessment Committee is working toward a different assessment for ELLs: "What we're attempting to do across the state is to test students against themselves. So all the different variables we've talked about ...we can't just say, they've been in Washington schools 3 or 4 years, then they should progress just like other kids. It's very individualized."

All administrators interviewed expressed dissatisfaction with aspects of the No Child Left Behind Act and WASL, although two mentioned positive benefits as well: the attention paid to "instructional technology" (a phrase from District D's superintendent) and the attention paid to ELLs' learning needs. Only one administrator (the state and federal programs director from District A) believed that ELLs should be included automatically in WASL from the start.

Reactions to the data report. The first interview question asked for administrators' reactions to the data report. Three themes emerged. One, many administrators were startled at the lack of progress by the ELLs' fourth and fifth years in district. Two, administrators were surprised at how much higher the performance was for

FEP students. Three, most administrators mentioned the many variables impacting ELL achievement and the need to consider those as well.

Superintendents in Districts A, C and D, along with the elementary education director in District D, voiced surprise and concern that students in their fourth and fifth year in the program did not appear to be doing well. District D's superintendent noted that "some kids did better in years 2-3 than they did in years 4-5"; his elementary education director noted "a plateau". The superintendent for District C found the data discouraging: "My theory is that in one or two years, chances [of meeting standard] are almost nil. And here, they're less than 10%. Even at 4-5 years, they're barely a quarter. That's telling." Yet the assistant superintendent in District E felt encouraged by the data report. He saw students making progress, slow but higher than previous years.

District B administrators, having been through extensive data analysis in their district already, were not surprised at how well the exited students performed. In District E, the assistant superintendent was surprised, but the special programs director had been gathering these data for quite some time: she called it the "dipstick approach to see that they have sustaining success". District B's assistant superintendent was worried that the visible success of the exited students on the bar graphs would lead to some conclusions that students should be exited sooner. She wondered if others would believe the bilingual program might be holding them back rather than realizing that the goals of the bilingual program had to be met first. In two instances, she proved to be right. District A's superintendent asked, "What would happen if you had them exit earlier? I mean, look at the math here. If I were a parent, and I saw that kind of dramatic improvement, I'd be wanting my kid to exit." (However, this superintendent was not aware of the exit criteria

for ELLs. It had to be explained to him. Perhaps because ELLs count for a small portion of this district, around 3%, he had not familiarized himself with regulations for this program.) District D's elementary education director also referred to the groups in the program for 4 and 5 years and stated he was concerned about the long-term effect of the bilingual program. However, he went on to say he thought the program had been plagued with lack of materials and weak instruction in the past. Several improvements have been implemented, and he expects WASL scores this year (2003-2004) to reflect that.

Two administrators insisted these data be shared with policymakers. One stated, "They need to see this. This study is objective. Here it is, this is what happens. Not in one, not in two years [ELLs meeting standard]...it can't happen." Similarly, another said, "Take what we already know and put it right in front of our faces. You can't argue with the data. This here is quite dramatic."

All administrators interviewed mentioned other variables besides time in district that were critical to ELL achievement. At least one administrator from each district mentioned both literacy in the home and poverty as key variables. However, administrators' way of viewing these variables differed. District E's director of special services mentioned, "A child who comes from a parent who isn't a reader and a writer or a limited reader and writer, and who doesn't experience the same kind of environment, it takes much longer for them to progress in literacy and math and everything the WASL expects". Superintendents from District A and D, along with the elementary education director from District D, noted that students from the Eastern European bloc appeared to achieve at a faster rate. The superintendent from District A stated: "I have a basic theory about all this. I think this is far more related to poverty than it is to language. The kids

we get from other countries who aren't poor, like kids from France and Germany, they're strong in their own language." He went on, "The black kids who don't speak academic English, I see similarities." Later, this superintendent said he felt many Mexican families were inhibited by their machismo culture and did not value an education or getting ahead, especially for their girls. District D's elementary education director mused: "For us, we have kids from the East European bloc countries that attend one school mostly, and they do pretty well after they've been here awhile, and their scores tend to continue to grow... Most of those kids have been to school... and I think most of them have come from... more affluent families in their home country, whereas our kids from Mexico, most of them are basically coming from poverty."

Other frequently mentioned variables were lack of former schooling and amount of comparable academic experience. As the District C superintendent explained, his typical new student tends to be from a migrant family, new to the country and on the average, 3 years behind academically. Administrators also referred to language development in the primary language and amount of English upon entry in their school district as critical variables. The home and its support for school was another cited variable. The following example was given as evidence of this: District E has a gifted program which accepts ELLs as well as English-only students. The director of special programs reported that these particular ELLs, "look a lot like a middle class white student in their environment. They have two parents, they've read with them at home in their home language, so they're much more advanced in their literacy and vocabulary skills".

Three additional variables of educational quality were brought up: teachers having too much to do, lack of professional development in appropriate instructional strategies

and lack of materials. Two administrators, both from districts with high migrant student counts, pointed out that absenteeism was an overlooked variable. What may look like an academic year is often not actually an academic year, due to many missed days or weeks, they explained.

WASL and ELL inclusion. Second, administrators were asked for recommendations for a more equitable way to include ELLs in the state assessment system. Their recommendations varied. Five administrators believed that a new assessment was needed, one that would show incremental growth. Four of them did not believe the Washington Language Proficiency Test was an appropriate measure, while one, the elementary education director from District A, was not familiar with the WLPT and so had no opinion. As one administrator stated, "It's not a format they're used to, it assesses a different kind of thinking than what we work on in class". Two others thought the WLPT was an appropriate measure for ELLs and that it should take the place of the WASL.

Five administrators recommended setting an established point of readiness prior to taking the WASL. District D's elementary education director said: "I think we need clear standards based on the WLPT about when the kid is ready to take the WASL in English. And it's got to come off a test like the WLPT, because otherwise districts are going to interpret it differently." District C's superintendent said, "Until we can guarantee that they are proficient in the language, they should not have to take it. It's not fair and it's not achievable." Three administrators thought students should actually exit the bilingual program before being required to take WASL. The assistant superintendent for District E stated, "Like your data show, kids that exit from the program are scoring at a

fairly high rate." The special programs director added, "The reason they've exited was because they've met a criteria of really high test scores."

While four administrators stressed that there should not be exemptions based on time alone, the superintendent of District D felt ELLs should have a 3-year exemption. "It's a political issue, not an educational issue. No matter how you cut that, politics is going to rule. If we could get 3 years, I think that's the best we can get. It's not necessarily an equal chance, but it's a better chance," he reasoned. District A's superintendent recommended a 1-year exemption. Three administrators believed that redesignated ELLs should continue to be held in the program. One stated that if programs are to be held accountable, then that should include all the students served by the bilingual program.

Only one administrator (the state and federal programs director for District A) felt all ELLs actually needed to be included in all state testing. She stated that it was important to have a baseline and any other testing information. District D's bilingual program director suggested that the students should take the tests the other students take but their scores should not be counted, rather than penalizing them or the other students in the district. She also suggested a system of expected yearly progress standards for ELLs, so teachers could have yardsticks to go by; these standards would meet up with WASL eventually.

Four administrators felt having the assessment in primary languages of ELLs would yield more valid information. However, District A's superintendent noted that with 26 languages in his district this year, it wasn't feasible. In Districts D and E, administrators wished they had WASL versions in Spanish and Russian for ELLs who

are stronger in those languages academically than English. (Both these districts have higher concentrations of these two language groups.)

While two administrators mentioned the positive effect of better teaching for all students, more administrators mentioned negative effects of testing ELLs with WASL. These included spending too much time on testing and having a harmful effect on student self-esteem. One administrator relayed that fourth grade ELLs took four major tests during the school year in their particular district. Another said, "The look on their faces, like 'I'm stupid'. Well, they're not." Another mentioned the blow to the students' self esteem and its effect on learning: "Another challenge is that after the test, their self esteem goes way down. It's like if someone gave us a test in Chinese and said, here, go take this. And the test takes such a big chunk of time. It causes just a big down shift in the learning momentum."

District A administrators, actively campaigning against upcoming WASL-like tests at further grades, argued that for the time and effort put into it, WASL had little useful information for educators. WASL does not help teachers know about the students they have in class, since the results aren't known until the next fall; also, WASL does not show individual progress. Further, their research shows the fourth grade WASL is not predictive of the seventh grade WASL.

Challenges with the fourth grade WASL and ELLs. The third question asked each administrator what they regarded as the biggest challenge with the WASL and ELL students. Most participants said there were many. While administrators had different perspectives and addressed different challenges, several mentioned the threat of not making AYP.

In District A, the elementary education director said, "It's just such a public spectacle. So threatening – the school's going to be taken over. Well, they're not going to do any better. And what does WASL really tell us? The biggest challenge is this: false assumptions." She did go on to add, "And another thing: I have never heard any clear direction about what we can do to help ELLs in the classroom perform on WASL. Teachers get all kinds of helpful suggestions and directives for mainstream kids." The district's state and federal programs director saw the biggest challenge as "finding ways to use the WASL as described without causing the district to fail in meeting AYP with that subgroup [ELLs]. That needs to be amended in some way." She also mentioned "the sheer number of assessments administered to ELL students" as a challenge (although, interestingly, she was the only administrator who stated that ELLs should always take the WASL along with the other students.)

In District B, the assistant superintendent felt dealing with all of the barriers to achievement were challenges. "There are kids barely coming in that we have to test right away. There are kids with no English and no former schooling. There are homes with no literacy. There are homes of extreme poverty. There are teachers with too much to do already." As far as WASL goes, she selected the automatic inclusion of ELL and Special Education scores in state results as the biggest obstacle. "It's ridiculous," she said.

The superintendent of District C felt the lack of student preparation was his biggest challenge. Getting his students to academic readiness was a hard fight; this is where his district has aimed their resources. He also saw attendance as a constant challenge. This district's literacy and assessment director saw one very large challenge

as finding bilingual people to teach the students. She added, "The biggest challenge is that we need an appropriate measure to show growth."

The superintendent for District D, one of the kindergarten readiness program creators, felt language development was a critical challenge. "No one knows how to deal with the students who are not proficient in either language. Not proficient in vocabulary or structure. And that can be a native English speaker, as well, who is not proficient in English. What can you do with them, what is most efficient, how can you get parents to understand that language development is so crucial?" He felt a connected challenge was "how quickly we can get the kids to learn English." The elementary education director for District D felt the challenge was finding the best materials and instructional programs. He also stated that "when people research this some more, there are some new answers coming that we don't have yet about how to best teach this group." The bilingual program director found the biggest challenge to be "the pressure for them to perform like native English speakers, and it's all AYP. Principals are all up tight, and the kids are nervous. But I think the biggest challenge is trying to give these kids more time. Dealing with AYP on the one hand and trying to help the kids relax even if they just got here from Mexico."

In District E, both administrators felt their main challenge was gaining time for their ELLs to achieve more proficiency. They also hoped to see a more appropriate assessment for ELLs. District E faces many challenges. As their assistant superintendent added: 'Language is [but] one of the issues that they're working with to be successful at school." Another challenge is that many students in District E do not qualify for the

bilingual program but struggle with low achievement (Appendix H contains a 1999-2000 graph comparing fourth grade WASL achievement of different demographic groups.)

Problems with the study. Participants mentioned two problems in the study. First, from the leap forward in WASL scores that exited students seem to take, it could be easy for people to misconstrue that the bilingual program held people back. One person pointed out to me that I would have to be very careful with my wording when I explained this part to people. Two people did in fact reach such a conclusion from studying the data, expressing concern that students were kept in the bilingual program too long, and I did have to explain about exit criteria and the exiting process, hopefully with care. Second, time in district by itself does not really tell about an individual student. Several administrators pointed that out to me. They wanted to be sure I knew that time alone would not bring achievement. So many variables influence how long it takes a student to reach academic proficiency, they pointed out. Among them are: length of schooling in home country, language development in native language, preparation for U.S. schools, literacy in the home, poverty level of the family and native culture, and quality of instruction. "It's a very individualized process", the special programs director in District E emphasized.

Qualitative component summary. This study served an informative purpose for many of the administrators. It was a surprise for most to see how slowly and unevenly achievement levels rose for ELLs in the subgroups representing district length of stay. The higher achievement levels of redesignated students were also surprising. Unexpectedly, the follow-up interviews also served informative purposes. For example, the superintendent in District A did not know about the exit process for ELLs in the state

bilingual program. Nor did he know about the WLPT. The elementary education director in District D was unaware of the variables that might help ELL performance at a high income low-incidence ELL school be higher than at other low income high-incidence ELL schools. These were subjects that we were able to discuss and clarify.

While district administrators brought different perspectives to the interviews, all administrators but one felt that ELLs could be assessed more appropriately by other means. These means included another incremental test to measure yearly progress, using the WLPT in place of the WASL rather than as an addition to the WASL, and translated versions of the WASL. One administrator favored a 3-year exemption; another suggested a 1-year exemption. Additionally, five people felt that a set of readiness criteria should be established for ELLs to meet before WASL inclusion, and two people felt that ELLs should entirely exit the bilingual program prior to taking WASL. Some referred to my study as evidence.

When administrators recommended more equitable ways to assess English language learners, giving them more time was frequently mentioned. Clearly, issues of time held import. Yet most administrators in the study made note of the many other variables influencing achievement as well. Poverty and literacy in the home were mentioned most frequently. Several others surfaced, such as amount of formal schooling, amount of English upon arrival, literacy in first language, and educational level of parents. However, in the two districts with less ELLs and with higher WASL achievement, three of the administrators mentioned country of origin and instruction as major variables. Each remarked on how the students from Bosnia and Eastern Europe seemed to learn more quickly than the students from Mexico. One superintendent went so

far as to say that parents from Mexico didn't value an education for their daughters. Further, the elementary education directors from both District A and E felt that if instruction were better organized, better WASL results would come from ELLs.

Summary

For these multidistrict samples over 3 testing years, 5 years in district was not enough time for most ELLs to meet standard on the state achievement tests. While dividing data into length of stay groups showed improvement as students stayed in district longer, improvement was often slight and uneven between the groups. In the areas of Math and Writing, improvement was less than in Reading and Listening. Yet scores of ELLs did not begin to approach scores of the mainstream fourth grade group in any area. Exited ELLs were another story. They met standard at a percentage rate close to mainstream students, sometimes surpassing them. In Math, however, there was a greater gap than in the other areas. When the ELL scores were removed from the grade level aggregate score, it usually resulted in about 10% more meeting standard.

In follow-up interviews with 12 administrators from the five districts, most were surprised by the lack of improved performance on WASL for ELLs in their district for 4 and 5 years. Also, the higher performance of the exited ELLs surprised them. All but one had recommendations for a more equitable way to include ELLs in WASL, including giving them an alternative test that shows incremental growth until they reach a point of readiness for WASL or until they exit the bilingual program.

Chapter 5

Discussion

"To the extent that we ensure that all students are included in educational accountability systems, the more likely we are to address how best to assess all students." (Thurlow & Liu, 2001, p.18).

"Until we can guarantee that they are proficient in the language, they should not have to take it [WASL]. It's not fair and it's not achievable." Superintendent, District C, February, 2004.

The No Child Left Behind Act presents educators and policymakers with a giant puzzle. All students must (and should) be included in state assessment systems, yet assessment should be appropriate for the learners (LaCelle-Peterson & Rivera, 1994). In the case of English language learners, however, state assessments under-represent their content knowledge (CCSSO, 2000; Abedi, 2004). Low ELL performance on state assessments is partly due to the high degree of linguistic complexity on high stakes tests (Abedi, 2004). Additionally, past research has shown that developing English academic language is a process that takes several years (Collier, 1992; Collier & Thomas, 2004; Cummins, 1981; De Avila, 1997; Hakuta, Butler & Witt, 2000; Ramirez, Yuen & Ramey, 1991; Thomas & Collier, 1997, 2002).

Description of findings

This study gives a more layered picture of ELL performance, from five districts during 3 testing years, on the fourth grade Washington's Assessment of Student Learning. My first three questions dealt with the WASL scores of ELL students over time. The first question asked how test performance on the WASL would differ for ELLs with 1 year, 2 years, 3 years, 4 years and 5 years in district. The second question inquired at what point ELL scores would become similar to native English-speaking scores. The third question asked how scores of former ELLs now reclassified as fluent English proficient compared with ELL and mainstream student scores. The study showed that, while test performance did improve in general with time in U.S. schools, it did not improve at a steady or incremental rate according to number of years in district. Improvement from 1 year in district to the next often appeared slight, while sometimes there was no improvement. At times, percentage meeting standard actually lessened for the group with a further year in district. In two areas, Listening and Reading, ELLs showed steadier progress over years in district. Yet they were still far below grade level even after 5 years in the district. On the other hand, students formerly in the bilingual program but now reclassified as FEP were performing at similar levels with the rest of fourth grade, especially in the area of Reading. Math and Writing remained more difficult for ELLs and FEPs, as well as for all students.

The study showed that, after 5 years of schooling in English, most ELLs in these districts failed to reach the desired standard for success on the fourth grade achievement test. By the fifth year most were still far behind their native English-speaking peers on WASL achievement, in all four test areas (Math, Reading, Writing and Listening). This

concurred with studies by Hakuta et al. (2000) and Ramirez et al. (1991), who both found that the entire span of elementary school was necessary for acquisition of academic English. One district participating in this study (District C) had already reached this conclusion; they had adopted a policy of not exiting ELLs from the bilingual program during the elementary years in order to keep all necessary support in place.

Further, findings on the differentiated achievement in the separate area tests seemed to confirm research by De Avila (1997), who found that elementary ELLs first improved in the area of listening and speaking, followed by reading and later by writing. He also demonstrated that this process took the entire span of elementary school. On the WASL, ELL performance follows De Avila's progression. Performance on Listening is higher than on the other three tests, followed by Reading. Writing as measured by the WASL is not strong at fourth grade, nor is Math. The Math test requires a great deal of writing, and the reading level on the Math test is higher than fourth grade level (Brickell & Lyon, 2003).

These findings also support previous research showing that ELLs need a minimum of 4 and possibly as much as 10 years in U.S. schools before they perform on English tests at grade level with native English-speaking peers (Collier, 1992; Cummins, 1981; De Avila, 1997; Hakuta, Butler & Witt, 2000; Ramirez, Yuen & Ramey, 1991; Thomas & Collier, 1997, 2002). Collier and Thomas' more recent research (2004) finds 6 to 8 years necessary.

My fourth question investigated how district scores would change with the ELL student scores removed. In the areas of Math and Reading, the percentages meeting standard rose by almost 10% in each of the testing years. District percentages passing on

the Writing and Listening tests also rose, but at a lesser rate, ranging from a 3% to 9% rise. In the event that Washington adopts an alternative way to assess ELLs, these data help forecast possible outcomes for participating districts.

My fifth and sixth questions were concerned with how the administrators from the districts in the study would react to the data. For the administrators from these five districts (with the exception of District B, where they had already pursued these data themselves), these were more detailed and useful data than what they previously possessed. Most of the administrators interviewed were surprised at both how low the ELLs in the fourth and fifth year in district performed and how high the exited students performed. Moreover, after reviewing the data, all administrators but one had suggestions for more equitable ways to include ELLs in the state assessment system. Several suggested a new incremental assessment, one that measured students against themselves, while two administrators thought the Washington Language Proficiency Test (already given yearly to all ELLs in the state to measure language development) would suffice as the yearly assessment. Students already have to take it; because it was developed for ELLs, it is more suitable. Thus it makes sense to use it in place of WASL until WASL readiness is reached by meeting a predetermined level of criteria, one reasoned. In fact, several administrators thought an established set of WASL readiness criteria should be established, and three administrators thought ELLs should exit the bilingual program completely before being required to take the WASL. One administrator thought yearly standards should be set for expected growth, designed to eventually meet up with the WASL. Another favored a 3-year exemption. Still another thought 1 year might be enough. Several mentioned translated WASL versions for ELLs.

Educational implications

This study provides insight into possible outcomes for these districts for the two newest provisions for the NCLBA (U.S. Department of Education, 2004). The first provision, allowing new ELLs an academic year before they need to be tested on their state's reading/ language arts assessment, will not be helpful in these districts. As the data show, there is not much difference in Reading achievement at 1 year or at 2 years. It is still very low at 2 years, 5%, 8% and 16% meeting standard for the 3 academic years of the study (1999-2000, 2000-2001, 2001-2002). By 5 years in district, only one-fourth of the fourth grade ELLs are meeting standard in Reading. Further, this figure has remained constant over the last 3 testing years. Clearly, 1 year to adjust to English schooling is not enough. Further, even 5 years are not enough for three-fourths of the students.

The second provision, to allow redesignated students to remain in the data pool for 2 more testing years with the ELL subgroup, holds more promise. This study showed that redesignated ELLs met standard at percentages much nearer to their grade level peers than the students still classified as ELLs; moreover, the redesignated students often performed on par with all students. For example, in Reading, 60%, 57%, and 66% of former ELLs met standard in the 3 consecutive academic years of the study; 59%, 58%, and 60% of all students met standard. In Writing, 31%, 36% and 50% of redesignated students met standard, while 37%, 41%, and 47% of all students met standard. Plainly, these former ELLs who tested out of the state bilingual program were able to achieve at a much higher level on WASL. Including the redesignated FEPs may well bring a degree of balance to the ELL aggregate score.

Similarly, this study examined the "revolving door" element (as one of the administrators in the study called it), by looking at the achievement of newcomers and students exiting the ELL program. New students moving in pull the scores down, and students exiting out take high-end scores with them (Abedi, 2004). In Washington, half of the state's ELL students moved in or out of the bilingual program within the 2001-2002 school year. For the students in this study, there was often no clear difference between students with 1 year or 2 years in their district. Even at 5 years, most students were not able to meet standards. In these districts, newcomer scores often did not appear that different from ELLs who had been in their district longer periods of time, especially in the areas of Math and Writing. However, exited students did perform at much higher levels. In these districts, a good deal of the high end scores did appear to have left with the FEPs.

This study underscored another major issue concerning statewide assessments and ELLs. For schools such as these with ELL subgroups, making AYP becomes a further challenge. The AYP design assumes that subgroups and schools start from the same point and progress at the same rate. Further, the ELL subgroup is expected to progress not only at the rate of other students, but at an even faster rate, since they must learn English skills at the same time that they are learning grade-level content. It is clear that in these districts, ELLs have far to go to be at the level of their native English-speaking peers. Moreover, their rate of progress is not steady or rapid. As Linn (2003) demonstrated, having low-performing subgroups in a school makes attainment of AYP unlikely. It is no surprise that many of the schools from these five districts are already in various phases of corrective action for failing to meet AYP.

A further issue is the validity of testing students in English while they are yet learning English. The high degree of English linguistic complexity on high stakes assessments makes them tests of English proficiency rather than tests of content knowledge (American Evaluation Association, 2000; August & Hakuta, 1997; Butler & Stevens, 1997; CCSSO, 2000; Hakuta, 2001; Holmes & Duron, 2000; LaCelle-Peterson & Rivera, 1994; Linn, 1998; National Research Council, 2000; Olson & Goldstein, 1997; O'Malley & Valdez Pierce, 1996; Rivera & Stansfield, 1998; Thurlow & Liu, 2001). For the former ELLs in this study who have been reclassified as fluent English proficient, WASL appears far less difficult, supporting the notion that with a higher degree of English proficiency, content knowledge on tests in English can be more readily demonstrated.

What is a better solution? Giving ELLs more time in district for WASL readiness might improve test performance. It could provide students more opportunity to break through the language barrier, untangle linguistic complexity and become familiar with testing structures. Yet this study shows that a set amount of years is not a sufficient gauge of readiness. In these participating districts, ELLs performed better at 4 and 5 years than they did at 1 and 2 years. However, it was not near the level of their English-speaking peers. Perhaps districts should report their data on ELL students as I did in this study, by number of years in program. The districts could be held accountable to increase the rate of learning evidenced by the students by the number of years in the ELL program (an idea also advanced by one of the administrators in this study).

The state already uses an English proficiency test, the Washington Language Proficiency Test (WLPT), to test all ELL students each spring. It shows incremental

growth in English language development. In a more ideal testing system, one possible solution could be for students to reach an established point of English academic readiness as measured by the state's English proficiency test before moving on to WASL. This idea was supported by two of the administrators interviewed for this study. With this solution, ELLs are both included in the assessment system but not measured as though they are English proficient.

Limitations of the study

There were several limitations to this study. First, as many of the participating administrators pointed out, it only looked at one part of the complex interaction of factors that effect student achievement. It did not look at family socio-economic status, educational level of the family, number of books in the home, or previous amount of schooling before enrolling in the district. The study did not account for amount of schooling in the first language, which researchers have found to be predictive of both the success and rate of acquisition of a new language (Alcala, 2000; Berman, 1997; Brisk, 1998; Collier & Thomas, 2004; Cummins, 1986, TESOL, 1997; Thomas & Collier, 1997). Next, the five participating districts were neighbors, but they were distinct. While combining them increased the sample size and enlarged the focus to a geographic area of the state, considering them separately may have gained different results than combining them. For example, during these 3 study years, District B was earning the second lowest WASL scores in the state, largely due to a highly migrant, high poverty population where many families came from a rural area of Mexico with limited schooling. District E is a large district with one of the largest Spanish-speaking student populations in the state; while its scores are improving, it struggles with low WASL achievement. District C was

seeing more growth than the others; the superintendent felt this was due to many grant-supported endeavors connected to the Bill and Melinda Gates Foundation award. Next, as a descriptive study, WASL achievement was measured by percentages meeting standards. These came from different sizes of groups, limiting comparability, since higher numbers of participants carry more stability and less variance. In 2001-2002, for example, 9 percent of 47 ELLs, with 1 year in district, has less statistical strength than the 33 percent of 328 ELLs with 5 years in district. Finally, although the total ELL achievement outcomes from these districts were somewhat close to the state, the data from a descriptive study cannot be generalized beyond these five participating districts, during these 3 testing years.

Areas for further study

Several possibilities remain for further study. Examining ELL scores on the seventh grade WASL by lengths of stay in these districts might add information of value. It might further support the need for 6 to 8 years that Collier and Thomas claim necessary for English academic achievement in their latest study (2004). Another area for study, suggested by two of the administrators, would be to follow students longitudinally, perhaps as cohorts who stay in district in comparison to newer students. (A study of this sort was done by District B, and a resulting graph can be found in Appendix G.) Studies that combine length of stay data with another critical variable (for example, students who enter U.S. schools with 2 or more years of formal schooling in L1) might provide further valuable insight. As one administrator suggested, another study might look at the students who had exited to see what they had in common; she hypothesized a common factor could be high quality schooling in their home country.

Conclusion

This study delved deeper into WASL achievement data to search for a relationship between time in district and achievement on Washington's fourth grade state test for ELLs. In a case study of five districts, over 3 testing years, ELL scores were disaggregated by their years in their district. The over-arching question for the study was, how long is long enough? Data showed that although scores improved with time, even 5 years in U.S. schools were not enough for most ELLs in these districts to meet standard in any of the four test areas (Math, Reading, Writing and Listening). While ELLs scored higher in Listening than the other three areas, the Listening test has been discontinued. In Reading, even with 5 years in a district, three fourths of the ELL students failed to meet standard. In Math and Writing, only 16 percent of the ELLs were able to meet standard with 5 years in district.

The findings from this study support previous length of time research on ELLs and academic achievement. The complexity and difficulty in acquiring academic competency in a new language makes it a lengthy process of several years. For students who were able to meet goals of the ELL program, who exited and were redesignated as fluent English proficient, achievement at levels closer to their native English-speaking peers were possible.

How long is long enough? In fourth grade, 5 years is all the schooling that most students have; it was not enough for most students in the ELL program. Further, acquisition of English is an individual process, affected by many variables (Collier, 1991; Collier & Thomas, 2004; Cummins, 1981, 1984; De Avila, 1997; Hakuta et al., 2000; Ramirez et al., 1991; Rivera & Stansfield, 2000; Thomas and Collier, 1997, 2002).

Assignation of a number of years does not seem enough by itself. A better answer might be, as long as it takes to reach fluent English proficient status. As Collier and Thomas (2004) concluded (p.1):

> Once ELL learners have reached full parity with native-English speakers, a curricular test in English should yield just as valid and reliable a score as it does for native-English speakers. But, while ELL learners are still closing the gap, a test score in English will underestimate their true achievement.

References

Abedi, J. (2004). The No Child Left Behind Act and English language learners: Assessment and accountability issues. *Educational Researcher, 33, 1,* 4-14.

Abedi, J., & Dietel, R. (2004). *Challenges in the No Child Left Behind Act for English language learners* (CRESST: Policy Brief, 7). Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing.

Abedi, J., Leon, S., & Mirocha, J. (2003). *Impact of students' language background on content-based assessment: Analyses of extent data* (CSE Tech. Rep. No. 603). Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing.

Abedi, J., & Lord, C. (2001). The language factor in mathematics tests. *Applied Measurement in Education, 14,* 219-234.

Abedi, J., Lord, C., & Plummer, J. (1997). *Language background as a variable in NAEP mathematics performance* (CSE Tech. Rep. No. 429). Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing.

Albus, D., Bielinski, J., Thurlow, M., & Liu, K. (2001). *The effect of a simplified English language dictionary on a reading test* (LEP Projects Report 1). Minneapolis: University of Minnesota, National Center on Educational Outcomes. Retrieved July 19, 2001 from: http://education.umn.edu/NCEO/OnlinePubs/LEP1.html

Alcala, A. (2000). A framework for developing an effective instructional program for limited English proficient students with limited former schooling. *Practical Assessment, Research & Evaluation, 7, 9.*

Allen, R. (2002). Acquiring English: Schools seek ways to strengthen language learning. *Association for Supervision and Curriculum Development (ASCD) Curriculum Update*, Fall 2002.

American Educational Research Association. (2004). English language learners: Boosting academic achievement. *Research Points: Essential Information for Education Policy, 2, 1*. Retrieved April 1, 2004 from http://www.aera.net/pubs/rp/RP_Winter04.pdf

American Evaluation Association. (2000). *Position statement concerning high-stakes testing in preK-12 education.* Fairhaven MA: AEA. Retrieved July 3, 2002 from http://www.eval.org/hst3.htm

August, D., & Hakuta, K. (1997). *Improving schooling for language minority children: A research agenda.* Washington DC: National Academy of Sciences, National Research Council, Board on Children, Youth and Families.

August, D., Hakuta, K.,& Pompa, D. (1994). *For all students: Limited English proficient students and Goals 2000* (NCBE FOCUS: Occasional Papers in Bilingual Education, 10). Washington D.C.: National Clearinghouse for Bilingual Education.

Baker, C., & Jones, S. P. (1998). *Encyclopedia of bilingualism and bilingual education.* Clevedon UK: Multilingual Matters.

Baker, K.A., & de Kanter, A.A. (1981). *Effectiveness of bilingual education: A review of literature*. Washington DC: Office of Planning, Budget and Evaluation, United States Department of Education.

Baker, K.A., & de Kanter, A.A. (1983). *Bilingual education.* Lexington MA: Lexington Books.

Beaumont, C., De Valenzuela, J.S., & Trumbull, E. (2002). Alternative assessment for transitional readers. *Bilingual Research Journal*, 26, *2*, 213- 240.

Bergeson, T. (2003, August). *Washington student achievement is up by any measure.* Press Release posted August 8, 2003, to http://www.k12.wa.us/Press/pressreleases/082803WASL.asp

Bergeson, T., & Heuschel, M.A. (2003, November). *A Washington State proposal to ensure successful implementation of No Child Left Behind.* Olympia: Office of the Superintendent of Public Instruction.

Berman, P. (1997). *Studies of education reform: School reform and student diversity.* Washington DC: U.S. Department of Education.

Bill & Melinda Gates Foundation. (n.d.) *Education: Washington State programs.* Retrieved November 20, 2002 from http://www.gatesfoundation.org/education /washingtonstateprograms/

Bloom, B., & Krathwohl, D. (1977). *Taxonomy of educational objectives: Handbook I: Cognitive domain.* White Plains NY: Longman.

Brickell, J., & Lyon, D. (2003). *Reliability, validity, and related issues pertaining to the WASL.* Federal Way WA: WEA Instructional Research. Retrieved June 27, 2003 from http://www.wa.nea.org/pubs/edpolicy/wasl/rel_validity_wasl2.pdf

Brisk, M.E. (1998). *Bilingual education: From compensatory to quality schooling.* Mahwah NJ: Lawrence Erlbaum Associates.

Bush, President George W. (2001, January). Moving limited English proficient students to English fluency, Title III. *No Child Left Behind.*

Butler, F.A., & Stevens, R. (1997). *Accommodation strategies for English language learners on large-scale assessments: Student characteristics and other characteristics.* (CSE Tech. Rep. No. 448). Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing.

Calderon, M. (1997). *Staff development in multilingual multicultural schools.* (Report No. EDO-UD-97-5). NY: ERIC Clearinghouse on Urban Education Digest (Digest No.124). Available: http://eric-web.tc.columbia.edu/digests/dig124.html

Chomsky, N. (1965). *Aspects of the Theory of Syntax.* Cambridge MA: MIT Press.

Chomsky, N. (1969). *The acquisition of syntax in children 5 to 10.* Cambridge MA: MIT Press.

Collier, V.P. (1992). A synthesis of studies examining long-term language minority student data on academic achievement. *Bilingual Research Journal, 16,* 187-212.

Collier, V.P., & Thomas, W.P. (2004). The astounding effectiveness of dual language education for all. *NABE Journal of Research and Practice, 2,* 1-20.

Commission on Student Learning. (1998). *Accountability system recommendations.* Olympia: Washington State Commission on Student Learning.

Council of Chief State School Officers. (1992). *Summary of recommendations and policy implications for improving the assessment and monitoring of students with limited English proficiency.* Washington DC: Council of Chief State School Officers.

Council of Chief State School Officers. (2000). *Ensuring accuracy in testing for English language learners.* Washington DC: Council of Chief State School Officers.

Council of Chief State School Officers. (2003). *Statewide educational accountability under NCLB: Central issues arising from an examination of state accountability workbooks and U.S. Department of Education reviews under the No Child Left Behind Act of 2001.* Washington DC: Council of Chief State School Officers.

Crawford, L.W. (1993). *Language and literacy learning in multicultural classrooms.* Boston: Allyn & Bacon.

Cummins, J. (1979). Linguistic interdependence and the educational development of bilingual children. *Harvard Educational Review, 49,* 222-251.

Cummins, J. (1981). The role of primary language development in promoting educational success for language minority students. In *Schooling and Language Minority Students.* Sacramento: California Department of Education, pp. 3-49.

Cummins, J. (1984). *Bilingualism and special education: Issues in assessment and pedagogy.* Clevedon England: Multilingual Matters.

Cummins, J. (1986). Empowering minority students: A framework for intervention. *Harvard Educational Review, 56,* 18-36.

Daggett, W.R. (2004). Proficiency levels vary across states, subjects and grades. *Views You Can Use, 3, 5.* Retrieved February 5, 2004 from http://www.icle.net/vyvu%20III%20%no%205.html

De Avila, E. (1997). *Setting expected gains for non and limited English proficient students* (NCBE Resource Collection Series 8). Washington DC: National Clearinghouse for Bilingual Education.

Erickson, F. (1986). Qualitative methods in research on teaching. In M.C. Wittrock (Ed.), *Handbook of research on teaching* (3rd ed., pp. 119-161). NY: Macmillan.

Escamilla, K., Mahon, E., Riley-Bernal, H., & Rutledge, D. (2003). High-stakes testing, Latinos, and English language learners: Lessons from Colorado. *Bilingual Research Journal, 27*, 25-49.

Gándara, P., & Merino, B. (1993). Measuring the outcomes of LEP programs: Test scores, exit rates, and other mythological data. *Educational Evaluation and Policy Analysis, 15, 3*, 320-338.

Glesne, C. (1999). *Becoming qualitative researchers.* (2nd ed.) NY: Longman.

Greene, J.P. (1998). *A meta-analysis of the effectiveness of bilingual education.* Austin TX: University of Texas, The Tomas Rivera Policy Institute.

Grimes, C. (2003, September 30). Law: All students will be tested. *Walla Walla Union-Bulletin*, pp. E4-E5.

Hakuta, K. (1986). *Mirror of language: The debate on bilingualism.* NY: Basic Books, Inc.

Hakuta, K. (2001). *The education of language minority students.* Testimony to the United States Commission on Civil Rights, April 13, 2001. Retrieved July 19, 2002 from http://www. stanford.edu/~hakuta/Docs/CivilRightsCommission.htm

Hakuta, K., Butler, Y.G., & Witt, D. (2000). *How long does it take English language learners to attain proficiency?* (Linguistic Minority Institute Policy Report). Santa Barbara: University of California Linguistic Minority Research Institute.

Hammond, B. (2004, April 11). Dual-language testing boon for state, pupils. *OregonLive.com.* Retrieved April 15, 2004 from http://www.oregonlive.com/printer/printer.ssf?/basenews/1081598749320860.xml

Hill, R.K., & DePascale, C.A. (2003). Reliability of No Child Left Behind accountability designs. *Educational Measurement: Issues and Practice, 22, 3,* 12-20.

Holmes, D., & Duron, S. (2000). *LEP students and high stakes assessment.* Washington DC: National Clearinghouse for Bilingual Education.

Houser, J. (1995). *Assessing students with disabilities and limited English proficiency* (Working Paper No. 95-13). Washington DC: National Center for Educational Statistics, Policy and Review Branch, Data Development Division.

Improving America's Schools Act (IASA) of 1994, Public Law 103-382-Oct. 24, 1994, Section IIII, Title I- Amendments to the Elementary and Secondary Act of 1965.

Joireman, J., & Abbot, M.L. (2001). *The relationships between the Iowa Test of Basic Skills and the Washington Assessment of Student Learning in the state of Washington* (Washington School Research Center: Technical Report #2). Lynnwood: Washington School Research Center.

Kindler, A.L. (2002). *Survey of the states' limited English proficient students and available educational programs and services 2000-2001 summary report.* Washington, DC: National Clearinghouse for English Language Acquisition & Language Instruction Educational Programs.

Kingsbury, G.G., Olson, A., Cronin, J., Hauser, C., & Houser, R. (2002). *The state of state standards: Research investigating proficiency levels in fourteen states.* Portland OR: Northwest Evaluation Association.

Kiplinger, V.L., Haug, C.A., & Abedi, J. (2000, June). *A math assessment should assess math, not reading: One state's approach to the problem.* Paper presented at the 30th National Conference on Large Scale Assessment, Snowbird, UT.

Krashen, S. (1996). *Under attack: The case against bilingual education.* Culver City CA: Language Education Associates.

LaCelle-Peterson, M.W., & Rivera, C. (1994). Is it real for all kids? A framework for equitable assessment policies for English language learners. *Harvard Educational Review, 64, 1,* 55-75.

Linn, R.L. (1998). *Standards-based accountability: Ten suggestions* (CRESST Policy Brief). Los Angeles: University of California, National Center for Research on Evalution, Standards and Student Testing.

Linn, R.L. (2003). *Requirements for measuring Adequate Yearly Progress* (Policy Brief 6). Los Angeles: University of California, National Center for Research on Evalution, Standards and Student Testing.

Linn, R.L., Baker, E.L., & Herman, J.L. (2002). *Minimum group size for measuring adequate yearly progress: The CRESST line.* Los Angeles: University of California, Center for the Study of Evaluation/ National Center for Research on Evaluation, Standards, and Student Testing.

Linquanti, R. (2001). *The redesignation dilemma: Challenges and choices in fostering meaningful accountability for English learners* (Policy Report no. 2001-1). Santa Barbara: University of California Linguistic Minority Research Institute.

Mabry, L., Poole, J., Redmond, L., & Schultz, A. (2003). Local impact of state testing in southwest Washington. *Education Policy Analysis Archives, 11, 22.* Retrieved July 18, 2003 from http://epaa.asu.edu/epaa/v11n22/

MacSwan, J. (2000). The Threshold Hypothesis, semilingualism, and other contributions to a deficit view of linguistic minorities. *Hispanic Journal of Behavioral Sciences, 22,1,* 3-45. Retrieved April 10 from http://www.public.asu.edu/~macswan/hjbs2000.pdf

MacSwan, J., Rolstad, K., & Glass, F.V. (2002). Do some school-age children have no language? Some problems of construct validity in the Pre-LAS Español. *Bilingual Research Journal, 26, 2,* 213-237.

MacSwan, J., & Rolstad, K. (2003). Linguistic diversity, schooling, and social class: Rethinking our conception of language proficiency in language minority education. In C.B. Paulson & R. Tucker (Eds.), *Essential Readings in Sociolinguistics* (pp. 329-340). Oxford: Blackwell.

McCall, M. (2004, January). *How do we learn language?* Paper presented at the 11th Annual Office of Superintendent of Public Instruction January Conference, Spokane, WA.

McDonnell, L., & Hill, P.T. (1993). *Newcomers in American schools: Meeting the educational needs of immigrant youth.* Rand. Retrieved July 1, 2001 from http://www.rand.org/publications

McKeon, D. (1994). When meeting "common" standards is uncommonly difficult. *Educational Leadership, 51, 8.* 45-49.

McLaughlin, B., & McLeod, B. (1996). Educating all our students: Improving education for children from culturally and linguistically diverse backgrounds. *Final Report of the National Center for Research on Cultural Diversity and Second Language Learning, 1,* June 1996.

Menken, K. (2000). *What are the critical issues in wide-scale assessment of English language learners?* (NCBE Issue Brief No. 6). Washington DC: National Clearinghouse for Bilingual Education.

MetriTech, Inc. (1999). *Language Proficiency Test Series: Assess English language learners and measure their progress.* Champaign IL: MetriTech/ IPI. Retrieved February 2, 2004 from http://www.metritech.com/Metritech/Products/lpts.htm

Miles, M.B., & Huberman, A.M. (1994). *Qualitative data analysis.* (2nd ed.) Thousand Oaks, CA: Sage Publications, Inc.

Miramontes, O.B., Nadeau, A., & Commins, N.L. (1997). *Restructuring schools for linguistic diversity: Linking decision-making to effective programs.* NY: Teachers College Press.

Moss, M., & Puma, M. (1995). *Prospects: The Congressionally mandated study of educational opportunity and growth: Language minority and limited English proficient students.* Washington DC: U.S. Department of Education.

Moon, T.R., Callahan, C.M., & Tomlinson, C.A. (2003) Effects of state testing programs on elementary schools with high concentration of student poverty – good new or bad news? *Current Issues in Education, 6.*

National Research Council. (2000). *Testing English language learners in U.S. schools: Report and workshop summary.* Committee on Educational Excellence and

Testing Equity. K. Hakuta & A. Beatty, Eds., Board on Testing and Assessment, Center for Education. Washington DC: National Academy Press.

No Child Left Behind Act of 2001, Public Law 107-110, H.R. 1. 107th Congress of the United States of America. (2002).

Northwest Evaluation Association (n.d.). *NWEA's Measures of Academic Progress (MAP) background.* Retrieved March 15, 2004, from http://www.nwea.org/newsroom/map_backgrounder.pdf

Office of Superintendent of Public Instruction. (2000a). *Bilingual education in Washington State: Annual review.* Olympia: Office of Superintendent of Public Instruction.

Office of Superintendent of Public Instruction. (2000b). *State laws regarding education reform, assessment, and accountability in Washington State.* Olympia: Office of Superintendent of Public Instruction.

Office of Superintendent of Public Instruction. (2000c). *Educating limited-English-proficient students in Washington State.* Olympia: Office of Superintendent of Public Instruction.

Office of Superintendent of Public Instruction. (2002a). *WASL test data.* [Data file]. Available from Office of Superintendent of Public Instruction Web site, http://www.k12.wa.us/edprofile/

Office of Superintendent of Public Instruction. (2002b). *Washington State educational profile.* [Data file]. Available from Office of Superintendent of Public Instruction Web site, http://www.k12.wa.us/edprofile/

Office of Superintendent of Public Instruction. (2002c). *Educating limited-English-proficient students in Washington State.* Olympia WA: Office of Superintendent of Public Instruction.

Office of Superintendent of Public Instruction. (2003a). *Educating limited-English-proficient students in Washington State: School Year 2001-2002.* Olympia WA: Office of Superintendent of Public Instruction.

Office of Superintendent of Public Instruction. (2003b). *Guidelines for participation and testing accommodations for special populations in state assessment programs.* Olympia: Office of Superintendent of Public Instruction.

Office of Superintendent of Public Instruction. (2003c). *Washington State report card.* [Data file]. Available from Office of Superintendent of Public Instruction Web site, http://www.k12.wa.us/edprofile/

Olson, J.F., & Goldstein, A.A. (1997). *The inclusion of students with disabilities and limited English proficient students in large-scale assessments.* Washington DC: National Center for Educational Statistics.

Olson, L. (2002). Quality Counts 2001: Finding the right mix. *Editorial Projects in Education, 20, 17,* 12-20. Retrieved July 14, 2002 from http://www.edweek.org/reports/qc01/articles/qc01story.cfm?slug=17intro.h20

O'Malley, J.M., & Valdez Pierce, L. (1996). Moving toward authentic assessment. In I.A. Heath & C.J. Serrano (Eds.), *Annual Editions 99/00: Teaching English as a second language* (p. 119-124). Guilford CT: Dushkin/ McGraw-Hill.

Orlich, D.C. (2003). An examination of the longitudinal effect of the Washington Assessment of Student Learning (WASL) on student achievement. *Education Policy*

Analysis Archives, 11, 18. Retrieved July 28, 2003 from http://eppaa.asu.edu/epaa/v11n18/

Pinker, S. (1994), *The language instinct: How the mind creates language*. New York: William Morrow and Company.

Ramirez, J.D., Yuen, S.D., & Ramey, D.R. (1991). *Executive summary final report: Longitudinal study of structured English immersion strategy, early-exit and late-exit transitional bilingual education programs for language minority children.* San Mateo CA: Aguirre International.

Rivera, C., & Stansfield, C.W. (1998). Leveling the playing field for English language learners. In R. Brandt (Ed.), *Assessing student learning: New rules, new realities*. Arlington VA: Educational Research Service.

Rivera, C., & Stansfield, C.W. (2000). *An analysis of state policies for the inclusion and accommodation of English language learners in state assessment programs during 1998-1999.* (Executive Summary). Washington DC: The George Washington University, Center for Equity and Excellence in Education.

Rivera, C., Vincent, C., Haffner, A., & LaCelle-Peterson, M. (1997). Statewide assessment programs: Policies and practices for the inclusion of L.E.P. students. *Practical Assessment, Research & Evaluation, 5.*

Rossell, C.H. (2000). Is one year enough? Can English learners attain proficiency in one-year immersion programs? *The ABC's of immersion: A teacher's guide.* Center for Equal Opportunity, pp. 6-8.

Rossell, C.H., & Baker, K. (1996). The educational effectiveness of bilingual education. *Review of Educational Research, 55, 3.*

Sanchez, R. (2000, April 30). Equal isn't fair in standardized testing. *Tri-City Herald*, pp. D3.

Shepard, L., Taylor, G., & Betebenner, D. (1998). *Inclusion of limited English proficient students in Rhode Island's grade 4 mathematics performance assessment.* (CSE Technical Report 486). Los Angeles: University of California, Center for the Study of Evaluation/ National Center for Research on Evaluation, Standards, and Student Testing.

Short, D.J., & Echeverria, J. (1999). *The Sheltered Instruction Observation Protocol: A tool for teacher-researcher collaboration and professional development.* Santa Cruz, CA: Center for Research on Education, Diversity & Excellence: (ERIC Document Reproduction Service No. ED434533)

Schulte, D.P., Schulte, L., Slate, J.R., & Brooks, G.D. (2002). A test-driven accountability system in Texas: Principals' voices. *International Electronic Journal for Leadership in Learning, 6, 7.* retrieved July 2, 2002 from http://www.ucalgary.ca/~iejll/volume6/slate.htm

Solomon, J., & Rhodes, N. (1995). *Conceptualizing academic language* (Research Report No. 15). Santa Cruz CA: the National Center for Research on Cultural Diversity and Second Language Learning.

Taylor, C.S. (1998). *Washington Assessment of Student Learning: Grade 4, 1998* (Technical Report). Olympia: Office of the Superintendent of Public Instruction.

Taylor, C.S., & Lee, Y. (2001). *Washington Assessment of Student Learning: Grade 4, 2000 (*Technical Report). Olympia: Office of Superintendent of Public Instruction.

Teachers of English to Speakers of Other Languages [TESOL]. (1997). *ESL standards for pre-K-12 students.* Alexandria VA: TESOL.

The Princeton Review. (2003). *Testing the testers 2003: An annual ranking of state accountability systems.* NY: Random House.

Thomas, W.P., & Collier, V. (1997). *School effectiveness for language minority students.* Washington DC: National Clearinghouse for Bilingual Education.

Thomas, W.P., & Collier, V. (2002). *A national study of school effectiveness for language minority students' long-term academic achievement.* Washington DC: Center for Research on Education, Diversity & Excellence. Retrieved February 10, 2003, from http://www.cal.org/crede/pubs/ResBrief10.htm

Thompson, M.S., DiCerbo, K.E., Mahoney, K., & MacSwan, J. (2002). ¿Exito en California? A validity critique of language program evaluations and analysis of English learner test scores. *Education Policy Analysis Archives, 10, 7.* Retrieved February 2, 2002 from http://cpaa/asu/cdu/cpaa/v10n7

Thurlow, M.L., & Liu, K.K. (2001). *State and district assessments as an avenue to equity and excellence for English language learners with disabilities* (LEP Projects Report 2). Minneapolic, MN: University of Minnesota, National Center on Educational Outcomes. Retrieved October 15, 2001 from http://education.umn.edu/NCEO/ OnlinePubs/LEP2.html

Thurlow, M.L., Liu, K.K., Quest, C., Thompson, S.J., Albus, D., & Anderson, M. (1998, Sept.). *Findings from research on accommodated statewide assessments for English language learners.* Paper presented at CRESST Conference, University of California, Los Angeles.

U.S. Department of Education. (2002). *Paige outlines adequate yearly progress provisions under No Child Left Behind.* Retrieved July 24 from http://www.ed.gov/ News/pressreleases/2002/ 07/07242002.html

U.S. Department of Education. (2004). *Secretary Paige announces new policies to help English language learners.* Retrieved February 22, 2004 from http://www.ed.gov/ news/pressreleases/2004/02/02192004.html

Voke, H. (2002). *What do we know about sanctions and rewards?* (ASCD InfoBrief, 31). Alexandria VA: Association for Supervision and Curriculum Development.

Walker, A., Shafer, J., & Iiamo, M. (2004). "Not in my classroom": Teacher attitudes towards English language learners in the mainstream classroom. *NABE Journal of Research and Practice, 2, 1,* 130-148.

Willig, A. (1985). A meta-analysis of selected studies of the effectiveness of bilingual education. *Review of Educational Research, 57, 3,* 269-317.

Wolcott, H.F. (1994). *Transforming qualitative data: Description, analysis, and interpretation.* Thousand Oaks, CA: Sage.

Wong-Fillmore, L. (1990). Now or later? Issues related to the early education of minority-group children. *Early childhood and family education: Analysis and recommendations of the Council of Chief State School Officers* (pp. 122-144). Orlando FL: Harcourt Brace Jovanovich.

Wright, W.E. (2002). The effects of high stakes testing in an inner-city elementary school: The curriculum, the teachers, and the English language learners. *Current Issues in Education, 5.* Tempe AZ: Arizona State University.

District Data Sheet for _____

Data requested by:

Claudia Olmstead Home: 943-2681 Work: 547-2441

Ph.D. student at U W in Education, Curriculum and Instruction

Dissertation focus: Fourth grade English Language Learners' performance on the W ASL and the relationship of scores to length of time in district, for school years 1999-2000, 2000-2001, 2001-2002.

Percentage of 4th grade bilingual program students with different lengths of stay who attain each WASL level (1 through 4)

1999-2000

WASL test:	With 1 year in district n=				With 2 years in district n=				With 3 years in district n=				With 4 years in district n=				With 5 years in district n=			
	W ASL levels:				WASL levels:				W ASL levels:				W ASL levels:				W ASL levels:			
	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
Math																				
Reading																				
Writing																				
Listening																				

1999-2000 continued

WASL test:	All students in program n=				Students who have exited program n=				All 4th graders in district n=				All non-bilingual program 4th graders n=			
	W ASL levels:				WASL levels:				W ASL levels:				W ASL levels:			
	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
Math																
Reading																
Writing																
Listening																

Percentage of 4th grade bilingual program students with different lengths of stay who attain each WASL level (1 through 4)

2000-2001

WASL test:	With 1 year in district n=				With 2 years in district n=				With 3 years in district n=				With 4 years in district n=				With 5 years in district n=			
	W ASL levels:				WASL levels:				W ASL levels:				W ASL levels:				W ASL levels:			
	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
Math																				
Reading																				
Writing																				
Listening																				

2000-2001 continued

WASL test:	All students in program n=				Students who have exited program n=				All 4th graders in district n=				All non-bilingual program 4th graders n=			
	W ASL levels:				WASL levels:				W ASL levels:				W ASL levels:			
	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
Math																
Reading																
Writing																
Listening																

Percentage of 4th grade bilingual program students with different lengths of stay who attain each WASL level (1 through 4)

2001-2002

WASL test:	With 1 year in district n=				With 2 years in district n=				With 3 years in district n=				With 4 years in district n=				With 5 years in district n=			
	WASL levels:				WASL levels:				WASL levels:				WASL levels:				WASL levels:			
	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
Math																				
Reading																				
Writing																				
Listening																				

2001-2002 continued

WASL test:	All students in program n=				Students who have exited program n=				All 4th graders in district n=				All non-bilingual program 4th graders n=			
	W ASL levels:				WASL levels:				W ASL levels:				W ASL levels:			
	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
Math																
Reading																
Writing																
Listening																

District Data Collection Form for **District A**

Number/ percentage of 4th grade bilingual program students with different lengths of stay who attain each WASL level (levels 1 –2 do not meet standard; levels 3-4 meet standard; writing and listening are only scored as does not meet/ meets standard), 1999-2000:

** NOTES: **Data for bilingual program students and students exited from the bilingual program were both provided by District A's Director of Elementary Education and by the district's Bilingual Program. Data for all students were collected from the Office of Superintendent of Public Instruction, Washington State Report Card: (http://reportcard.ospi.k12.wa.us/Reports/WASL). Figures for all non-bilingual program 4th graders were calculated by the researcher.**

WASL test:	With 1 year in district n= 6 [nt: 5- included w/ others for uniformity w/ other districts]				With 2 years in district n= 5				With 3 years in district n= 1				With 4 years in district n= 0				With 5 years in district n= 2			
	WASL levels:				WASL levels:				WASL levels:				WASL levels:				WASL levels:			
	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
Math	6 100	0 00	0 00	0 00	3 60.0	1 20.0	1 20.0	0 00	1 100	0 00	0 00	0 00	0 00	0 00	0 00	0 00	1 50	1 50	0 00	0 00
Reading	4 66.7	2 33.3	0 00	0 00	0 00	3 60.0	2 40.0	0 00	1 100	0 00	0 00	0 00	0 00	0 00	0 00	0 00	1 50	1 50	0 00	0 00
Writing	6 100		0 00		4 80.0		1 20.0		1 100		0 00		0 00		0 00		2 100		0 00	
Listening	6 100		0 00		3 60.0		2 40.0		1 100		0 00		0 00		0 00		2 100		0 00	

1999-2000 continued:

WASL Test:	All students in program n= 14				Students who have exited program n= 6				All 4th graders in district math n= 702 nt:9 (1.3%) reading n= 700 nt:12 (1.7%) writing n= 697 [nt:14] listening n= 703 [nt:10]				All non-bilingual program 4th graders math n= 688 nt: 9 (1.3%) reading n= 686 nt:12(1.7%) writing n= 683 [nt:14] listening n= 689 [nt:10]			
	WASL levels:				WASL levels:				WASL levels:				WASL levels:			
	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
Math	11 78.6	2 14.3	1 7.1	0 00	1 16.7	2 33.3	1 16.7	2 33.3	170 24.2	202 28.8	182 25.9	139 19.8	159 23.1	200 29.1	181 26.3	139 20.2
Reading	6 42.9	6 42.9	2 14.3	0 00	0 00	0 00	5 83.3	1 16.7	23 3.2	170 24.3	330 47.2	165 23.6	17 2.5	164 23.9	328 47.8	165 24.1
Writing	13 92.9		1 7.1		1 16.7		5 83.3		366 52.5		331 47.5		353 51.7		330 48.3	
Listening	12 85.7		2 14.3		0 00		6 100		191 27.1		512 72.9		179 26.0		510 74.0	

Number/ percentage of 4th grade bilingual program students with different lengths of stay who attain each WASL level (levels 1 –2 do not meet standard; levels 3-4 meet standard; writing and listening are only scored as does not meet/ meets standard), 2000-2001:

WASL test:	With 1 year in district n= 6				With 2 years in district n= 4				With 3 years in district n= 4 writing nt: 1				With 4 years in district n= 0				With 5 years in district n= 3 writing nt= 1			
	WASL levels:				WASL levels:				WASL levels:				WASL levels:				WASL levels:			
	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
Math	5 83.3	0 00	0 00	1 16.7	4 100	0 00	0 00	0 00	4 100	0 00	0 00	0 00	0 00	0 00	0 00	0 00	2 66.7	1 33.3	0 00	0 00
Reading	2 33.3	2 33.3	2 33.3	0 00	3 75	1 25	0 00	0 00	1 25.0	3 75.0	0 00	0 00	0 00	0 00	0 00	0 00	0 00	2 66.7	1 33.3	0 00
Writing	5 83.3		1 16.7		4 100		0 00		4 100		0 00		0 00		0 00		2 66.7		1 33.3	
Listening	5 83.3		1 16.7		3 75.0		1 25.0		3 75.0		1 25.0		0 00		0 00		2 66.7		1 33.3	

2000-2001 continued:

WASL Test:	All students in program n= 17 writing nt:2				Students who have exited program n= 4				All 4th graders in district math n= 704 nt: 3 (.04%) reading n= 702 nt:5 (.07%) writing n= 694 [nt:12] listening n=704 [nt:3]				All non-bilingual program 4th graders math n= 687 nt:3 (.04%) reading n= 685 nt:5 (.07%) writing n= 677 [nt:12] listening n= 687 [nt:3]			
	WASL levels:				WASL levels:				WASL levels:				WASL levels:			
	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
Math	15 88.2	1 5.9	0 00	1 5.9	0 00	1 25.0	1 25.0	2 50.0	159 22.6	183 26.0	172 24.3	187 26.6	144 21.0	182 26.5	172 25.0	186 27.1
Reading	6 35.3	8 47.1	3 17.6	0 00	0 00	0 00	3 75.0	1 25.0	32 4.5	155 22.1	313 44.6	197 28.1	26 3.8	147 21.5	310 45.3	197 28.8
Writing	13 76.5		4 23.5		0 00		4 100		351 50.6		343 49.4		338 49.9		339 50.1	
Listening	13 76.5		4 23.5		1 25.0		3 75.0		142 20.2		562 79.8		129 18.8		558 81.2	

Number/ percentage of 4th grade bilingual program students with different lengths of stay who attain each WASL level (levels 1 –2 do not meet standard; levels 3-4 meet standard; writing and listening are only scored as does not meet/ meets standard), 2001-2002:

WASL test:	With 1 year in district n= 1				With 2 years in district n= 2				With 3 years in district n= 9				With 4 years in district n= 1				With 5 years in district n= 2			
	WASL levels:				WASL levels:				WASL levels:				WASL levels:				WASL levels:			
	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
Math	1 100	0 00	0 00	0 00	0 00	2 100	0 00	0 00	5 55.5	2 22.2	2 22.2	0 00	1 100	0 00	0 00	0 00	2 100	0 00	0 00	0 00
Reading	0 00	1 100	0 00	0 00	0 00	1 50	0 00	1 50	5 55.5	1 11.1	3 33.3	0 00	0 00	1 100	0 00	0 00	2 100	0 00	0 00	0 00
Writing	1 100		0 00		2 100		0 00		5 55.5		2 22.2		1 100		0 00		2 100		0 00	
Listening	0 00		1 100		1 50		1 50		4 44.4		5 55.5		1 100		0 00		0 00		2 100	

2001-2002 continued:

WASL Test:	All students in program: n= 15				Students who have exited program n= 7				All 4th graders in district math n= 711 nt: 7 (1.0%) reading n= 711 nt:6 (.8%) writing n= 700 [nt:16] listening n= 715 [nt:4]				All non-bilingual program 4th graders math n= 696 nt:7 (1.0%) reading n= 696 nt:6 (.9%) writing n= 685 [nt:16] listening n= 700 [nt: 4]			
	WASL levels:				WASL levels:				WASL levels:				WASL levels:			
	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
Math	9 60.0	4 26.7	2 13.3	0 00	1 14.3	2 28.6	4 57.1	0 00	129 18.1	203 28.6	202 28.4	170 24.0	120 17.2	199 28.6	200 28.7	170 24.4
Reading	5 33.3	6 40.0	3 20.0	1 6.7	0 00	2 28.6	5 71.4	0 00	33 4.7	203 28.5	291 40.9	178 25.1	28 4.0	197 28.3	288 41.4	177 25.4
Writing	13 86.7		2 13.3		2 28.6		5 71.4		344 49.2		356 50.8		331 48.5		354 51.5	
Listening	8 53.3		7 46.7		5 71.4		2 28.6		237 33.2		478 66.8		229 32.7		471 67.3	

District Data Sheet for **District B**

Number/ Percentage of 4th grade bilingual program students with different lengths of stay who attain each WASL level (1 through 4), 1999-2000:

** NOTES: **All data for District B were provided by the district.**

WASL test:	With 1 year in district n= 4				With 2 years in district n= 7				With 3 years in district n= 3				With 4 years in district n= 6				With 5 years in district n= 35			
	WASL levels:				WASL levels:				WASL levels:				WASL levels:				WASL levels:			
	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
Math	4 100	0 00	0 00	0 00	7 100	0 00	0 00	0 00	3 100	0 00	0 00	0 00	5 83.3	1 16.7	0 00	0 00	32 91.4	3 8.6	0 00	0 00
Reading	4 100	0 00	0 00	0 00	5 71.4	2 28.6	0 00	0 00	3 100	0 00	0 00	0 00	4 66.7	1 16.7	1 16.7	0 00	17 48.6	15 42.9	3 8.6	0 00
Writing	4 100		0 00		7 100		0 00		3 100		0 00		6 100		0 00		35 100		0 00	
Listening	4 100		0 00		7 100		0 00		3 100		0 00		4 66.7		2 33.3		24 68.6		11 31.4	

1999-2000 continued:

WASL Test:	**All students in program** n= 55				**Students who have exited program** n= 12				**All 4th graders in district** n= 101				**All non-bilingual program 4th graders** n= 46			
	WASL levels:				WASL levels:				WASL levels:				WASL levels:			
	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
Math	51 92.7	4 7.3	0 00	0 00	4 33.3	6 50.0	1 8.3	1 8.3	71 70.3	15 14.9	10 9.9	5 4.9	20 43.5	11 23.9	10 21.7	5 10.9
Reading	33 60.0	18 32.7	4 7.3	0 00	0 00	6 50.0	6 50.0	0 00	39 38.6	33 32.7	22 21.8	7 6.9	6 13.0	15 32.6	18 39.1	7 15.2
Writing	55 100		0 00		8 66.7		4 33.3		85 84.2		16 15.8		30 65.2		16 34.8	
Listening	42 76.4		13 23.6		7 58.3		5 41.7		62 61.4		39 38.6		20 43.5		26 56.5	

Number/ percentage of 4th grade bilingual program students attaining each WASL level (1 through 4) in year 2000-2001:

WASL test:	With 1 year in district n= 2				With 2 years in district n= 9				With 3 years in district n= 11				With 4 years in district n= 15				With 5 years in district n= 50			
	WASL levels:				WASL levels:				WASL levels:				WASL levels:				WASL levels:			
	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
Math	2 100	0 00	0 00	0 00	8 88.9	1 11.1	0 00	0 00	11 100	0 00	0 00	0 00	14 93.3	1 6.7	0 00	0 00	44 88.0	6 12.0	0 00	0 00
Reading	1 50.0	1 50.0	0 00	0 00	7 77.8	2 22.2	0 00	0 00	6 54.5	5 45.5	0 00	0 00	5 33.3	10 66.7	0 00	0 00	27 54.0	22 44.0	1 2.0	0 00
Writing	2 100		0 00		9 100		0 00		11 100		0 00		15 100		0 00		47 94.0		3 6.0	
Listening	2 100		0 00		8 88.9		1 11.1		10 90.9		1 9.1		11 73.3		4 26.7		36 72.0		14 28.0	

2000-2001 continued:

WASL Test:	**All students in program** n= 87				**Students who have exited program** n= 31				**All 4th graders in district** n= 135				**All non-bilingual program 4th graders** n= 48			
	WASL levels:				WASL levels:				WASL levels:				WASL levels:			
	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
Math	79 90.8	8 9.2	0 00	0 00	18 58.1	6 19.4	5 16.1	2 6.5	107 79.3	16 11.9	5 3.7	7 5.2	28 58.3	8 16.7	5 10.4	7 14.6
Reading	46 52.9	40 46.0	1 1.1	0 00	2 6.5	17 54.8	11 35.5	1 3.2	51 37.8	63 46.7	15 11.1	6 4.4	5 10.4	23 47.9	14 29.2	6 12.5
Writing	84 96.6		3 3.4		19 61.3		12 38.7		112 83.0		23 17.0		28 58.3		20 41.7	
Listening	67 77.0		20 23.0		20 64.5		11 35.5		92 68.1		43 31.9		25 52.1		23 47.9	

Number/ percentage of 4th grade bilingual program students attaining each WASL level (1 through 4) in year **2001-2002**:

WASL test:	**With 1 year in district** n= 3				**With 2 years in district** n= 5				**With 3 years in district** n= 8				**With 4 years in district** n= 6				**With 5 years in district** n= 48			
	WASL levels:				WASL levels:				WASL levels:				WASL levels:				WASL levels:			
	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
Math	3 100	0 00	0 00	0 00	5 100	0 00	0 00	0 00	7 87.5	0 00	1 12.5	0 00	6 100	0 00	0 00	0 00	40 83.3	6 12.5	2 4.2	0 00
Reading	1 33.3	2 66.7	0 00	0 00	5 100	0 00	0 00	0 00	6 75.0	2 25.0	0 00	0 00	6 100	0 00	0 00	0 00	29 60.4	14 29.2	5 10.4	0 00
Writing	3 100		0 00		5 100		0 00		8 100		0 00		6 100		0 00		47 97.95		1 2.1	
Listening	3 100		0 00		5 100		0 00		7 87.5		1 12.5		5 83.3		1 16.7		37 77.1		11 22.9	

2001-2002 continued:

WASL Test:	**All students in program** n= 70				**Students who have exited program** n= 31				**All 4th graders in district** n= 123				**All non-bilingual program 4th graders** n= 53			
	WASL levels:				WASL levels:				WASL levels:				WASL levels:			
	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
Math	61 87.1	6 8.6	3 4.3	0 00	8 25.8	14 45.2	6 19.4	3 9.7	79 64.2	24 19.5	11 8.9	9 7.3	18 34.0	18 34.0	8 15.1	9 17.0
Reading	47 67.1	18 25.7	5 7.1	0 00	4 12.9	13 41.9	11 35.5	3 9.7	58 47.2	36 29.3	21 17.1	8 6.5	11 20.8	18 34.0	16 30.2	8 15.1
Writing	69 98.6		1 1.4		21 67.7		10 32.3		104 84.6		19 15.4		35 66.0		18 34.0	
Listening	57 81.4		13 18.6		18 58.1		13 41.9		86 69.9		37 30.1		29 54.7		24 45.3	

District Data Collection Form for **District C**

Number/ percentage of 4th grade bilingual program students with different lengths of stay who attain each WASL level (levels 1 –2 do not meet standard; levels 3-4 meet standard; writing and listening are only scored as does not meet/ meets standard), **1999-2000**:

** NOTES: **All student data were provided by District C's Literacy/ Assessment Director and the elementary principal. Figures for all non-bilingual program 4th graders were calculated by the researcher.**

WASL test:	With 1 year in district n= 5				With 2 years in district n= 0				With 3 years in district n= 3				With 4 years in district n= 2				With 5 years in district n= 18			
	WASL levels:				WASL levels:				WASL levels:				WASL levels:				WASL levels:			
	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
Math	5 100	0 00	0 00	0 00	0 00	0 00	0 00	0 00	2 66.7	0 00	1 33.3	0 00	2 100	0 00	0 00	0 00	8 44.4	6 33.3	3 16.7	1 5.6
Reading	3 60	2 40	0 00	0 00	0 00	0 00	0 00	0 00	0 00	2 66.7	1 33.3	0 00	2 100	0 00	0 00	0 00	0 00	11 61.1	6 33.3	1 5.6
Writing	5 100		0 00		0 00		0 00		2 66.7		1 33.3		2 100		0 00		10 55.6		8 44.4	
Listening	4 80		1 20		0 00		0 00		2 66.7		1 33.3		2 100		0 00		11 61.1		7 38.9	

1999-2000 continued:

WASL Test:	All students in program n=28 (nt: 1)				Students who have exited program * n= 0 (policy of district: no program exit given for K-5)				All 4th graders in district math n= 83 reading n= 84 nt:2 (2.4%) writing n= 84 [nt: 2] listening n= 84				All non-bilingual program 4th graders math n= 55 reading n= 56 nt:2 (3.6%) writing n= 56 [nt: 3] listening n= 56			
	WASL levels:				WASL levels:				WASL levels:				WASL levels:			
	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
Math	17 60.7	6 21.4	4 14.3	1 3.6	0	0	0	0	53 63.9	19 22.9	10 12.0	1 1.2	36 65.4	13 23.6	6 10.9	0 00
Reading	5 17.9	15 53.6	7 25.0	1 3.6	0	0	0	0	6 7.1	45 53.6	28 33.3	3 3.6	1 1.8	30 53.6	21 37.5	2 3.6
Writing	19 67.9		9 32.1		0		0		56 66.7		28 33.3		37 66.1		19 33.9	
Listening	19 67.9		9 32.1		0		0		44 52.4		40 47.6		25 44.6		31 55.4	

Number/ percentage of 4th grade bilingual program students with different lengths of stay who attain each WASL level (levels 1 –2 do not meet standard; levels 3-4 meet standard; writing and listening are only scored as does not meet/ meets standard), **2000-2001**:

WASL test:	With 1 year in district n= 3				With 2 years in district n= 1				With 3 years in district n= 5				With 4 years in district n= 12				With 5 years in district n= 15			
	WASL levels:				WASL levels:				WASL levels:				WASL levels:				WASL levels:			
	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
Math n=36	0 00	3 100	0 00	0 00	0 00	100 100	0 00	0 00	0 00	4 80.0	1 20.0	0 00	0 00	10 83.3	2 16.7	0 00	0 00	12 80	3 20	0 00
Reading n=36	0 00	2 66.7	1 33.3	0 00	0 00	100 100	0 00	0 00	0 00	4 80.0	1 20.0	0 00	0 00	9 75	3 25	0 00	0 00	9 60	6 40	0 00
Writing n=36	3 100		0 00		1 100		0 00		5 100		0 00		8 66.7		4 33.3		9 60		6 40	
Listening n=36	2 66.7		1 33.3		0 00		1 100		3 60		2 40		4 33.3		8 66.7		9 60		6 40	

2000-2001 continued:

WASL Test:	All students in program n= 36				Students who have exited program n= 0 (policy of district: no program exit in K-5)				All 4th graders in district n= 83 math nt: 1 (1.2%) reading nt: 2 (2.4%) writing [nt: 2]				All non-bilingual program 4th graders n=~~45~~ ~~48~~ 47 math nt: 2 reading nt: 2			
	WASL levels:				WASL levels:				WASL levels:				WASL levels:			
	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
Math	30 83.3	0 00	6 16.7	0 00	0	0	0	0	52 62.7	19 22.9	10 12.0	1 1.2	22 46.8	19 40.4	4 8.5	1 2.1
Reading	0	25 69.4	11 30.6	0	0	0	0	0	5 6.0	45 54.2	28 33.7	3 3.6	5 10.6	20 42.6	17 36.2	3 6.4
Writing	26 72.2		10 27.8		0		0		55 66.3		28 33.7		29 61.7		18 38.3	
Listening	18 50.0		18 50.0		0		0		43 51.8		40 48.2		25 53.2		22 46.8	

Number/ percentage of 4th grade bilingual program students with different lengths of stay who attain each WASL level (levels 1 –2 do not meet standard; levels 3-4 meet standard; writing and listening are only scored as does not meet/ meets standard), **2001-2002**:

WASL test:	With 1 year in district n= 3				With 2 years in district n= 3				With 3 years in district n= 3				With 4 years in district n= 0				With 5 years in district n= 16			
	WASL levels:				WASL levels:				WASL levels:				WASL levels:				WASL levels:			
	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
Math n= 25	3 100	0 00	0 00	0 00	3 100	0 00	0 00	0 00	3 100	0 00	0 00	0 00	0	0	0	0	10 62.5	5 31.2	1 6.3	0 00
Reading n= 25	3 100	0 00	0 00	0 00	2 66.7	1 33.3	0 00	0 00	1 33.3	2 66.7	0 00	0 00	0	0	0	0	4 25.0	9 56.2	3 18.8	0 00
Writing n= 25	3 100		0 00		3 100		0 00		3 100		0 00		0		0		11 68.8		5 31.2	
Listening n= 25	3 100		0 00		3 100		0 00		3 100		0 00		0		0		12 75.0		4 25.0	

2001-2002 continued:

WASL Test:	All students in program n= 25				Students who have exited program n= 0 (policy of district: no program exit in K-5)				All 4th graders in district n= 75 math nt: 2 (2.6%) writing n= 75 [nt: 1]				All non-bilingual program 4th graders n= 50 math nt: 2 (4.0%) writing n= ~~47 [nt: 1]~~ 50			
	WASL levels:				WASL levels:				WASL levels:				WASL levels:			
	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
Math	19 76.0	5 20.0	1 4.0	0 00	0	0	0	0	25 33.8	22 29.9	19 24.7	7 9.1	6 12.0	17 34.0	18 36.0	7 14.0
Reading	10 40.0	12 48.0	3 12.0	0 00	0	0	0	0	10 13.3	37 49.3	22 29.3	6 8.0	0 00	25 50.0	19 38.0	6 12.0
Writing	20 80.0		5 20.0		0		0		42 56.0		33 44.0		22 44.0		28 56.0	
Listening	21 84.0		4 16.0		0		0		43 57.3		32 42.7		22 44.0		28 56.0	

District Data Sheet for **District D**

Number/ percentage of 4th grade bilingual program students with different lengths of stay who attain each WASL level (1 through 4), 1999-2000:

** NOTES: Data for bilingual program students and students exited from the bilingual program were provided by District D's bilingual program coordinator. Data for all students were collected from the Office of Superintendent of Public Instruction, Washington State Report Card (http://reportcard.ospi.k12.wa.us/Reports/WASL). Figures for all non-bilingual program 4th graders were calculated by the researcher.

WASL test:	**With 1 year in district** n= 14 [writing nt: 1]				**With 2 years in district** n= 18				**With 3 years in district** n=9 math nt: 1 (11.1%)				**With 4 years in district** n=10				**With 5 years in district** n= 39 reading nt: 1(2.6%)			
	WASL levels:				WASL levels:				WASL levels:				WASL levels:				WASL levels:			
	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
Math	10 71.4	2 14.3	2 14.3	0 00	15 83.3	2 11.1	1 5.6	0 00	5 55.6	2 22.2	1 11.1	0 00	8 80.0	2 20.0	0 00	0 00	28 71.8	9 23.0	1 2.6	1 2.6
Reading	4 28.6	8 57.1	2 14.3	0 00	4 22.2	13 72.2	1 5.6	0 00	1 11.1	8 88.9	0 00	0 00	1 10.0	7 70.0	2 20.0	0 00	9 23.1	21 53.8	7 17.9	1 2.6
Writing	13 92.9		1 7.1		17 94.4		1 5.6		8 88.9		1 11.1		9 90.0		1 10.0		35 89.7		4 10.3	
Listening	9 64.3		5 35.7		15 83.3		3 16.7		6 66.7		3 33.3		8 80.0		2 20.0		28 71.8		11 28.2	

1999-2000 continued:

WASL Test:	**All students in program** n= 90 math nt: 1(1.1%) reading nt: 1(1.1%) [writing nt: 1]				**Students who have exited program** n= 12				**All 4th graders in district** math n= 1036 nt:14 (1.4%) reading n= 1032 nt: 19 (1.8%) writing n= 1031 [nt: 22] listening n= 1037 [nt:16]				**All non-bilingual program 4th graders** math n= 947 nt: 14 (1.5%) reading n= 943 nt: 19 (2.1%) writing n= 942 [nt: 22] listening n= 947 [nt:16]			
	WASL levels:				WASL levels:				WASL levels:				WASL levels:			
	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
Math	67 74.4	16 17.8	5 5.6	1 1.1	6 50.0	6 50.0	0 00	0 00	308 29.7	254 24.5	236 22.8	224 21.6	241 25.4	238 25.1	231 24.4	223 23.6
Reading	19 21.1	57 63.3	12 13.3	1 1.1	0 00	6 50.0	6 50.0	0 00	39 3.8	266 25.8	450 43.6	258 25.	20 2.1	209 22.2	438 46.4	257 27.3
Writing	82 91.1		8 8.9		8 66.7		4 33.3		593 57.5		438 42.5		512 54.4		430 45.6	
Listening	66 73.3		24 26.7		5 41.7		7 58.3		358 34.5		679 65.5		292 30.8		655 69.2	

Percentage of 4th grade bilingual program students attaining each WASL level (1 through 4) in year 2000-2001:

WASL test:	With 1 year in district n=22 math nt: 1(4.5%) [writing nt: 2]				With 2 years in district n= 14				With 3 years in district n= 19 [writing nt: 1]				With 4 years in district n= 11				With 5 years in district n= 49 math nt: 2 (4.1%) reading nt: 2 (4.1%) [writing nt: 3]			
	WASL levels:				WASL levels:				WASL levels:				WASL levels:				WASL levels:			
	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
Math	19 86.4	2 9.1	0 00	0 00	13 92.9	1 7.1	0 00	0 00	14 73.7	2 10.5	3 15.8	0 00	9 81.8	2 18.2	0 00	0 00	35 71.4	8 16.3	4 8.2	0 00
Reading	5 22.7	14 63.6	3 13.6	0 00	3 21.4	11 78.6	0 00	0 00	4 21.1	11 57.9	3 15.8	1 5.3	0 00	5 45.5	6 54.5	0 00	10 20.4	23 46.9	13 26.5	1 2.0
Writing	20 90.9		2 9.1		13 92.9		1 7.1		15 78.9		4 21.1		8 72.7		3 27.3		46 93.9		3 6.1	
Listening	18 81.8		4 18.2		11 78.6		3 21.4		12 63.2		7 36.8		4 36.4		7 63.6		24 49.0		25 51.0	

2000-2001 continued:

WASL Test:	**All students in program** n= 115 math nt: 3 (2.6%) reading nt: 2 (1.7%) [writing nt: 6]				**Students who have exited program** n= 8				**All 4th graders in district** math n= 1008 nt: 7 (0.7%) reading n= 1006 nt: 6 (0.6%) writing n= 994 [nt: 18] listening n=1017 [nt: 7]				**All non-bilingual program 4th graders** math n= 893 nt: 7 (.8%) reading n= 893 nt:6 (.7%) writing n= 879 [nt: 18] listening n= 902 [nt:7]			
	WASL levels:				WASL levels:				WASL levels:				WASL levels:			
	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
Math	89 77.4	16 13.9	6 5.2	1 0.9	4 50.0	2 25.0	1 12.5	1 12.5	273 27.1	253 25.1	218 21.6	257 25.5	184 20.5	237 26.5	212 23.7	256 28.6
Reading	22 19.1	64 55.7	25 21.7	2 1.7	0 00	2 25.0	5 62.5	1 12.5	54 5.4	263 26.1	450 44.7	233 23.2	32 3.6	199 22.3	425 47.6	231 25.9
Writing	102 88.7		13 11.3		6 75.0		2 25.0		501 50.4		493 49.6		399 45.4		480 54.6	
Listening	69 60.0		46 40.0		1 12.5		7 87.5		255 25.1		762 74.9		186 20.6		716 79.4	

Percentage of 4th grade bilingual program students attaining each WASL level (1 through 4) in year **2001-2002:**

WASL test:	**With 1 year in district** n= 9				**With 2 years in district** n= 20 math nt:1 (5%) reading nt:1 (5%) [writing nt: 2]				**With 3 years in district** n= 13				**With 4 years in district** n= 21 [writing nt: 19]				**With 5 years in district** n= 53 [writing nt: 1]			
	WASL levels:				WASL levels:				WASL levels:				WASL levels:				WASL levels:			
	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
Math	3 33.3	5 55.6	1 11.1	0 00	14 70.0	3 15.0	1 5.0	1 5.0	9 69.2	3 23.1	1 7.7	0 00	7 33.3	8 38.1	5 23.8	1 4.8	26 49.1	14 26.4	12 22.6	1 1.9
Reading	2 22.2	5 55.6	2 22.2	0 00	6 30.0	9 45.0	3 15.0	1 5.0	2 15.4	8 61.5	3 23.1	0 00	3 14.3	12 57.1	5 23.8	1 4.8	6 11.3	27 50.9	19 35.9	1 1.9
Writing	7 77.8		2 22.2		17 85.0		3 15.0		13 100		0 00		16 76.2		5 23.8		41 77.4		12 22.6	
Listening	8 88.9		1 11.1		18 90.0		2 10.0		11 84.6		2 15.4		12 57.1		9 42.9		33 62.3		20 37.7	

2001-2002 continued:

WASL Test:	**All students in program** n= 116 math nt: 1 (.9%) writing [nt: 5]				**Students who have exited program** n= 7 nt: 1 (14.3%)				**All 4th graders in district** math n= 1071 nt: 5 (0.5%) reading n= 1065 nt: 9 (0.8%) writing n= 1058 [nt: 17] listening n=1077 [nt:5]				**All non-bilingual program 4th graders** math n= 955 nt: 5 (0..5%) reading n= 949 nt: 9 (0.9%) writing n= 942 [nt: 17] listening n= 961 [nt: 5]			
	WASL levels:				WASL levels:				WASL levels:				WASL levels:			
	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
Math	59 50.9	33 28.4	20 17.2	3 2.6	1 14.3	1 14.3	1 14.3	3 42.9	178 16.6	228 21.3	280 26.1	380 35.5	119 12.5	195 20.4	260 27.2	377 39.5
Reading	19 16.4	61 52.6	32 27.6	4 3.5	0 00	1 14.3	3 42.9	2 28.6	44 4.1	247 23.2	396 37.2	369 34.7	25 2.6	186 19.6	364 38.4	365 38.5
Writing	94 81.0		22 19.0		3 42.9		4 57.1		466 44.0		592 56.0		372 39.4		570 60.6	
Listening	82 70.7		34 29.3		2 28.6		5 71.4		363 33.7		714 66.3		281 29.2		680 70.8	

District Data Collection Form for **District E**

Percentage of 4th grade bilingual program students with different lengths of stay who attain each WASL level (levels 1 –2 do not meet standard; levels 3-4 meet standard; writing and listening are only scored as does not meet/ meets standard), **1999-2000**:

** **NOTES:** Data for bilingual program students and students exited from the bilingual program were provided by District E's Director of Special Programs. Data for all students were collected from the Office of Superintendent of Public Instruction, Washington State Report Card (**http://reportcard.ospi.k12.wa.us/Reports/WASL**). Figures for all non-bilingual program 4th graders were calculated by the researcher.

(n=331) WASL test:	With 1 year in district n= 9				With 2 years in district n= 25				With 3 years in district n= 23				With 4 years in district n= 24				With 5 years in district n= 250			
	WASL levels:				WASL levels:				WASL levels:				WASL levels:				WASL levels:			
	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
Math	9 100		0 00		23 92.0		2 8.0		22 95.7		1 4.3		22 91.7		2 8.33		230 92		20 8	
Reading	9 100		0 00		25 100		0 00		21 91.3		2 8.7		17 71.0		7 29.0		193 77.2		57 22.7	
Writing	8 88.9		1 11.1		23 92.0		2 8.0		22 95.7		1 4.3		22 91.7		2 8.33		218 87.2		32 12.8	
Listening	7 77.8		2 22.2		20 80.0		5 20.0		20 87.0		3 13.0		15 62.5		9 37.5		178 71.2		72 28.8	

1999-2000 continued:

WASL Test:	All students in program n= 331	Students exited from program n= 67	All 4th graders in district math n= 707 nt: 4 (.06%) reading n= 706 nt: 8 (1.1%) writing n= 681 [nt: 31] listening n= 715 [nt:5]	All non-bilingual program 4th graders math n= 376 nt:4 (1.1%) reading n= 375 nt:8 (2.1%) writing n= 350 [nt: 31] listening n= 384 [nt:5]
	WASL levels:	WASL levels:	WASL levels:	WASL levels:
	1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4
Math	306 92.4 25 7.6	48 71.7 19 28.3	377 53.3 149 21.1 110 15.6 67 9.4 526 74.4 177 25.0	220 58.5 152 40.4
Reading	265 80.1 66 19.9	27 40.3 40 59.7	119 16.9 297 42.0 220 31.1 62 8.8 416 58.9 282 39.9	151 40.3 216 57.6
Writing	293 88.5 38 11.5	50 74.6 17 25.4	525 77.1 156 22.9	232 66.3 118 33.7
Listening	240 72.5 91 27.5	32 47.8 35 52.2	395 55.3 320 44.7	155 40.4 229 59.6

Percentage of 4th grade bilingual program students with different lengths of stay who attain each WASL level (levels 1 –2 do not meet standard; levels 3-4 meet standard; writing and listening are only scored as does not meet/ meets standard), 2000-2001:

(n= 380) WASL test:	With 1 year in district n= 22				With 2 years in district n= 35				With 3 years in district n= 29				With 4 years in district n= 40				With 5 years in district n= 254			
	WASL levels:				WASL levels:				WASL levels:				WASL levels:				WASL levels:			
	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
Math	22 100		0 00		30 85.7		5 14.3		28 96.6		1 3.4		33 82.5		7 17.5		226 89.0		28 11.0	
Reading	22 100		0 00		30 85.7		5 14.3		23 79.3		6 20.7		28 70.0		12 30		188 74.0		66 26.0	
Writing	22 100		0 0		35 100		0 00		24 82.8		5 17.2		36 90.0		4 10.0		213 84.0		41 16.0	
Listening	19 86.4		3 13.6		27 77.1		8 22.9		21 72.4		8 27.6		24 60.0		16 40.0		158 62.2		96 37.8	

2000-2001 continued:

WASL Test:	All students in program n= 380				Students exited from program n= 70				All 4th graders in district math n= 702 nt: 1(0.1%) reading n= 700 nt:5 (.7%) writing n= 660 [nt:42] listening n= 705 [nt: 4]				All non-bilingual program 4th graders math n= 322 nt: 1 (0.3) reading n= 320 nt:5 (1.6%) writing n= 280 [nt: 42] listening n= 325 [nt: 4]			
	WASL levels:				WASL levels:				WASL levels:				WASL levels:			
	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
Math	~~158~~ ~~92.6~~ 339 89.2%		~~28~~ ~~7.4~~ 41 10.8%		49 70.0		21 30.0		362 51.6	156 22.2	101 14.4	82 11.7	~~185~~ 179 512 55.6%		~~156~~ 142 48.4 44.1%	
									518 73.8		183 26.1					
Reading	~~275~~ 291 ~~73.3~~ 76.6		~~105~~ 89 ~~17.7~~ 23.4		28 40.0		42 60.0		132 18.9	283 40.4	206 29.4	74 10.6	~~148~~ 124 ~~127.8~~ 38.8%		~~175~~ 191 ~~59.7~~ 59.7%	
									415 59.3		280 40.0					
Writing	~~347~~ 330 ~~86.8~~ 86.8		~~39~~ 50 ~~13.2~~ 13.2		47 67.1		23 22.9		497 75.3		163 24.7		~~136~~ 167 ~~59.7~~ 59.7		~~124~~ 113 44.3 40.4%	
Listening	~~256~~ 249 ~~65.5~~ 65.5		~~129~~ 131 ~~34.5~~ 34.5		24 34.3		46 65.7		350 49.6		355 50.4		~~141~~ 101 ~~31.1~~ 31.1		~~251~~ 224 ~~68.9~~ 68.9	

Percentage of 4th grade bilingual program students with different lengths of stay who attain each WASL level (levels 1 –2 do not meet standard; levels 3-4 meet standard; writing and listening are only scored as does not meet/ meets standard), **2001-2002:**

(n=362) WASL test:	With 1 year in district n= 31				With 2 years in district n= 28				With 3 years in district n= 38				With 4 years in district n= 56				With 5 years in district n= 209			
	WASL levels:				WASL levels:				WASL levels:				WASL levels:				WASL levels:			
	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
Math	30 96.8		1 3.2		25 89.3		3 10.7		32 84.2		6 15.8		44 78.6		12 21.4		173 82.8		36 17.2	
Reading	30 96.8		1 3.2		24 85.7		4 14.3		30 78.9		8 21.1		38 67.9		18 32.1		153 73.2		56 26.8	
Writing	29 93.5		2 6.5		26 92.9		2 7.1		35 92.1		3 7.9		44 78.6		12 21.4		173 82.8		36 17.2	
Listening	29 93.5		2 6.5		24 85.7		4 14.3		25 65.8		13 34.2		43 76.8		13 23.2		138 66.0		71 34.0	

2001-2002 continued:

WASL Test:	All students in program n= 362				Students exited from program n= 71				All 4th graders in district math n= 730 nt:12 (1.6%) reading n= 729 nt:10(1.6%) writing n= 711 [nt: 31] listening n= 741 [nt: 8]				All non-bilingual program 4th graders math n= 368 nt:12 (3.3%) reading n= 367 nt:10 (2.7%) writing n= 349 [nt: 31] listening n= 379 [nt: 8]			
	WASL levels:				WASL levels:				WASL levels:				WASL levels:			
	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
Math	~~321~~ 304 ~~88.7~~ 84.0		~~61~~ 58 ~~11.3~~ 16.0		36 50.7		35 49.3		310 42.5	152 20.9	152 20.9	104 14.2	~~141~~ 158 38.3 42.9		~~215~~ 198 58.4 53.8	
									462 ~~654~~ 63.3		256 35.1					
Reading	~~286~~ 275 ~~78.9~~ 76.0		~~76~~ 87 ~~21.1~~ 24.0		19 26.8		52 73.2		125 17.1	266 36.5	206 28.3	122 16.8	~~105~~ 116 28.6		~~186~~ 241 65.7	
									391 53.6		328 ~~45.1~~ 45.0					
Writing	307 ~~330~~ ~~88.1~~ 84.8		55 ~~42~~ ~~11.9~~ 15.2		32 45.0		39 55.0		468 65.8		243 34.2		~~143~~ ~~153~~ ~~42.9~~ 161		~~201~~ 188 ~~57.6~~	
Listening	259 ~~271~~ ~~750~~ 71.6		~~97~~ 103 ~~23.0~~ 28.5		22 31.0		49 69.0		384 51.8		357 48.2		~~173~~ 125 ~~29.3~~		~~256~~ 254 ~~70.2~~	

Appendix C

District Data Summary Report

Data Summary Report:

Fourth Grade English Language Learners Scores on the WASL

And Lengths of Stay In-District

By Claudia Olmstead

As part of a dissertation study for the College of Education,

University of Washington

February 2, 2004

Overview of study

A wealth of past research has shown a clear relationship between time in U.S. schools and grade-level academic performance for English language learners (Collier, 1992; Cummins, 1981; De Avila, 1997; Hakuta, Butler & Witt, 2000; Krashen, 1996; Ramirez, Yuen & Ramey, 1991; Thomas & Collier, 1997, 2002). Studies have demonstrated that at least four to seven years are necessary for most students. New federal legislation (the *No Child Left Behind Act of 2001*) requires including English language learners in annual state assessments in spite of limited English proficiency. Aligned with state achievement standards, these new assessments have influenced what gets taught in the classroom and how.

To observe a possible relationship between length of time in school district and achievement on this state's fourth grade WASL (Washington Assessment of Student Learning), descriptive testing data were collected for English language learners in bilingual programs in five school districts. Students were divided into the following subgroups, representing their length of time in district:

- Students with one year or less in district
- Students with one to two years in district
- Students with two to three years in district
- Students with three to four years in district
- Students with four to five or more years in district (these would be students enrolled in the district continually since kindergarten)

Also, another subgroup of students- those exited from their district's bilingual program- was added:

- Students previously in the district's bilingual program but exited by state criteria from the program

Study participants

Five neighboring districts agreed to participate in this study. The five districts were selected for both their availability and their variety; an attempt was made to represent a demographic range (see table below), like that found in the state's districts' bilingual programs. No attempt was made to compare districts or models of instruction; models of delivery included bilingual instruction, ESL pull-out, ESL inclusion, content ESL, and a combination of models. Data were combined to build a larger sample. To further increase the sample size, data were collected for three consecutive school years (1999-2000, 2000-2001, 2001-2002). For purposes of the study, data were combined and described for each year, and then data for all three years were combined and described as one large group.

In all cases, people from each district were helpful in gathering and providing data for the study. It often proved challenging to pinpoint how long individual students had been in the district and to identify and track down students who had exited. Only one of the five districts had already disaggregated their data in this manner.

District	Number of students in district	Language groups in bilingual program	Percentage of students in bilingual program
A	8,970	43% Ukrainian and Russian, 20% Spanish, 37% mix of 18 other languages	3%
B	1,600	Spanish	57%
C	821	Spanish	29%
D	13,210	84% Spanish, 6% Bosnian, 10% mix of 18 other languages	9.5%
E	8,185	97% Spanish, 3% mix of 11 other languages	40.6%

Note. Data for 2000-2001 from the Office of Superintendent of Public Instruction (2002). Educating limited-English-proficient students in Washington State. Olympia WA: Office of Superintendent of Public Instruction.

Study questions

The questions for this study included:

1. How does WASL test performance for ELLs differ at 1, 2, 3, 4, and 5 years in district?
2. At what point do ELL scores become similar to the scores of their native English-speaking peers?
3. How do test data for students who have exited the bilingual program compare to those for students still in the program and with other mainstream students?
4. If the districts' ELL scores, grouped together and unreflective of different lengths of time in district, are reported separately from the mainstream, how do the mainstream scores change?
5. What will administrators from the five districts think about the data?
6. After reviewing the data, what do administrators from the five districts recommend as an equitable way to include ELLs in the state assessment system?

Method

Each district was provided with a packet of tables to use for recording the test data for each of the three years. Data included: number of students per subgroup, percentage within each subgroup achieving each level for each WASL test (Math, Reading, Writing and Listening). The achievement levels attained by all fourth graders in the districts were also recorded. (Sometimes this information came from the districts, and sometimes it came from the OSPI Report Card website.) For a final comparison, ELL scores were removed from the fourth grade scores.

Completed data were then transcribed and combined onto an Excel spreadsheet. Using Excel, sums and percentages were then calculated for subgroups of ELL students from all five districts passing each WASL test. Tables for each year were prepared along with a table combining all three years. Tables summarizing the data with bar graphs were then prepared, treating each

WASL test (Math, Reading, Writing, and Listening) separately. These summary tables and bar graphs for each WASL test follow the report, along with the more detailed tables for each year.

For the follow-up portion of this study, district administrators in each of the five districts are provided with a copy of the data report (in most cases, these administrators include the superintendent, the elementary education director, and the bilingual program administrator). Administrators are asked to attend a meeting with the researcher, lasting no longer than 30 minutes. At this time, responses to a questionnaire will be sought and recorded (a copy of the questionnaire is included in this report). The researcher will identify themes that emerge from the responses to the questions and include them in the study.

Findings

First, it must be cautioned that because this is a descriptive study, with subgroups and samples of different sizes, groups cannot be directly compared.

How does WASL test performance for ELLs differ at 1, 2, 3, 4 and 5 years in district? It does improve in general with time, but it does not improve year by year, and it remains far below 4th grade level. The first and second year in district can almost be lumped together: performance is low, and one year is not necessarily higher than the other. In the third year performance often looks better, especially in Reading and Listening. In the fourth and fifth year, achievement is generally higher, but not always. Again, the fourth and fifth year group can almost be lumped together: the fourth year group often outperforms the fifth year group. When at its highest, bilingual program performance still remains about 30 percentage points behind all fourth graders. Reading and Listening show the most improvement in general, and they show the most improvement between the subgroups- but again, it is somewhat sporadic. Math shows the lowest performance of all.

This difference of achievement in the content areas seems to echo findings of DeAvila (1997), who studied progression of language development of ELLs who entered U.S. schools in

kindergarten. Listening performance improved steadily within the first two years, but reading achievement did not begin to rise until into the third year. Writing (which is also involved in WASL mathematics responses) was the last area to improve, and it began to show an improvement in the fourth year of schooling.

At what point do ELL scores begin to approach mainstream scores? How do test data for students who have exited the bilingual program compare to those for students still in the program and with other mainstream students? These questions have the same answer: Only when students have met the goals of the bilingual program do their scores begin to compare with all fourth grade students. This makes sense, since to exit the bilingual program students must demonstrate a functional level of academic competence in English. In Washington, exit criteria is correlated to the 35th percentile of grade level achievement testing. To qualify for the bilingual program, therefore, means to be functioning well below grade level.

How do district scores change when ELL scores are removed? It ranges from –3% (Writing, 1999-2000) to 12% (Reading, 2000-2001). The average is by nine percentage points.

Finally, if a relationship is found between district length of stay and WASL achievement of fourth grade ELLs, how might district administrators use this information? This question will be answered in the follow-up interview. (Note: when this study was originally designed, states had some discretion as to when and how to include ELLs in state assessments. That has changed, however, with NCLB. ELLs must now be included in state testing from the start.)

Resources:

Collier, V.P. (1992). A synthesis of studies examining long-term language minority student data on academic achievement. *Bilingual Research Journal, 16*, 187-212.

Cummins, J. (1981). The role of primary language development in promoting educational success for language minority students. In *Schooling and Language Minority Students.* Sacramento, CA: California Department of Education, pp. 3-49.

De Avila, E. (1997). *Setting expected gains for non and limited English proficient students.* Washington DC: National Clearinghouse for Bilingual Education. NCBE Resource Collection Series 8.

Hakuta, K.; Butler, Y.G. & Witt, D. (2000). *How long does it take English language learners to attain proficiency?* Santa Barbara CA: University of California Linguistic Minority Institute Policy Report.

Krashen, S.D. (1996). *Under attack: The case against bilingual education.* Culver City CA: Language Education Associates.

Ramirez, J.D., Yuen, S.D., & Ramey, D.R. (1991). *Executive summary final report: Longitudinal study of structured English immersion strategy, early-exit and late-exit transitional bilingual education programs for language minority children.* San Mateo CA: Aguirre International.

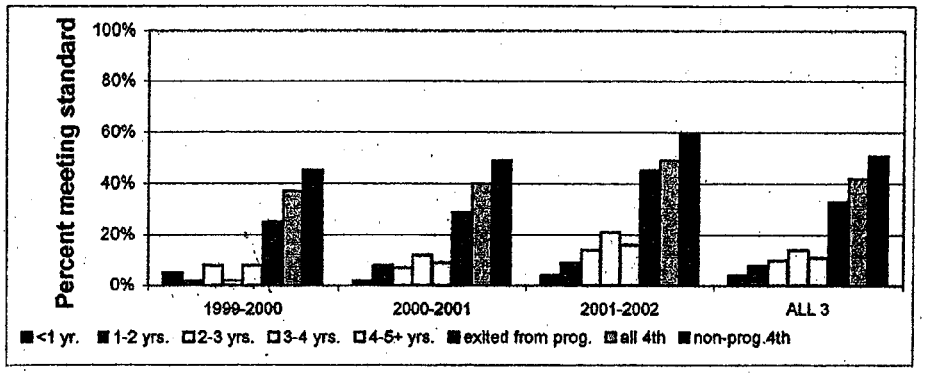
Thomas, W.P. & Collier, V. (1997). *School effectiveness for language minority students.* Washington DC: National Clearinghouse for Bilingual Education.

Thomas, W.P. & Collier, V. (2002). *A national study of school effectiveness for language minority students' long-term academic achievement.* Washington DC: Center for Research on Education, Diversity & Excellence. Retrieved February 10, 2003, from http://www.cal.org/crede/pubs/ResBrief10.htm

Grade 4 WASL Percent Meeting Standard *(refer to tables for n's)*

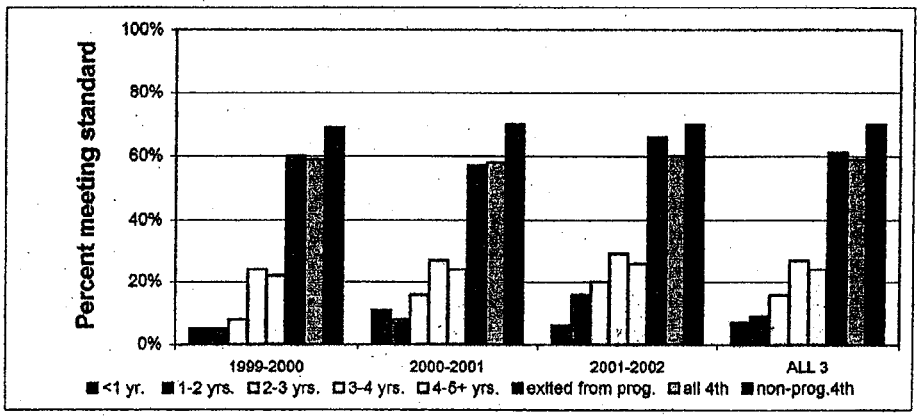
Math

TIME IN DIST:	1999-2000	2000-2001	2001-2002	ALL 3
<1 yr.	5%	2%	4%	4%
1-2 yrs.	2%	8%	9%	8%
2-3 yrs.	8%	7%	14%	10%
3-4 yrs.	2%	12%	21%	14%
4-5+ yrs.	8%	9%	16%	11%
exited from prog.	25%	29%	45%	33%
all 4th	37%	40%	49%	42%
non-prog.4th	45%	49%	59%	51%



Reading

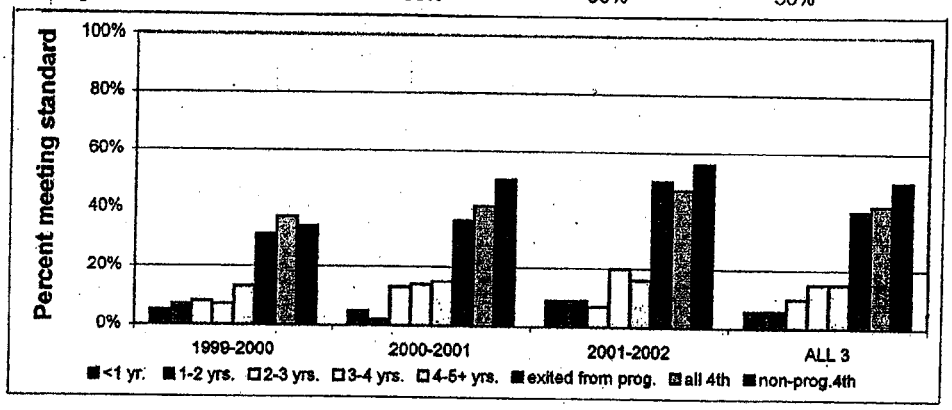
TIME IN DIST:	1999-2000	2000-2001	2001-2002	ALL 3
<1 yr.	5%	11%	6%	7%
1-2 yrs.	5%	8%	16%	9%
2-3 yrs.	8%	16%	20%	16%
3-4 yrs.	24%	27%	29%	27%
4-5+ yrs.	22%	24%	26%	24%
exited from prog.	60%	57%	66%	61%
all 4th	59%	58%	60%	59%
non-prog.4th	69%	70%	70%	70%



Grade 4 WASL Percent Meeting Standard *(refer to tables for n's)*

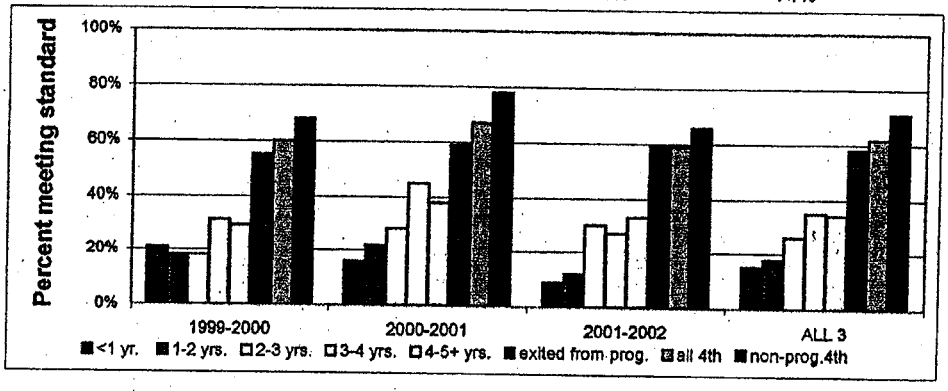
Writing

TIME IN DIST:	1999-2000	2000-2001	2001-2002	ALL 3
<1 yr.	5%	5%	9%	6%
1-2 yrs.	7%	2%	9%	6%
2-3 yrs.	8%	13%	7%	10%
3-4 yrs.	7%	14%	20%	15%
4-5+ yrs.	13%	15%	16%	15%
exited from prog.	31%	36%	50%	40%
all 4th	37%	41%	47%	42%
non-prog.4th	34%	50%	56%	50%



Listening

TIME IN DIST:	1999-2000	2000-2001	2001-2002	ALL 3
<1 yr.	21%	16%	9%	15%
1-2 yrs.	18%	22%	12%	18%
2-3 yrs.	18%	28%	30%	26%
3-4 yrs.	31%	45%	27%	35%
4-5+ yrs.	29%	38%	33%	34%
exited from prog.	55%	59%	59%	58%
all 4th	60%	67%	59%	62%
non-prog.4th	68%	78%	66%	71%



1999-2000			Percent of ELLs in district bilingual programs meeting standard					
Notes: red n's are ELLs, blue n's are regular 4th; n/a: District C- no exit policy for K-5								
MATH	**<1 year**	**1-2 years**	**2-3 years**	**3-4 years**	**4-5+ yrs**	**exited**	**all 4th**	**non-prog.**
Dist/ ELL #								
A 14	0	1	0	0	0	3	321	320
B 55	0	0	0	0	0	2	15	15
C 28	0	0	1	0	4	n/a	11	6
D 90	2	1	1	0	2	0	460	454
E 331	0	2	1	2	20	19	177	152
ttl	**2**	**4**	**3**	**2**	**26**	**24**	**984**	**947**
518 n	38	55	39	42	344	97	2,629	2,111
%	**5%**	**2%**	**8%**	**2%**	**8%**	**25%**	**37%**	**45%**
READING	**<1 year**	**1-2 years**	**2-3 years**	**3-4 years**	**4-5+ yrs**	**exited**	**all 4th**	**non-prog.**
Dist/ ELL #								
A 14	0	2	0	0	0	6	495	493
B 55	0	0	0	1	3	6	29	25
C 28	0	0	1	0	7	n/a	31	23
D 90	2	1	0	2	8	6	708	695
E 331	0	0	2	7	57	40	282	216
ttl	**2**	**3**	**3**	**10**	**75**	**58**	**1,545**	**1,452**
518 n	38	55	39	42	344	97	2,623	2,105
%	**5%**	**5%**	**8%**	**24%**	**22%**	**60%**	**59%**	**69%**
WRITING	**<1 year**	**1-2 years**	**2-3 years**	**3-4 years**	**4-5+ yrs**	**exited**	**all 4th**	**non-prog.**
Dist/ ELL #								
A 14	0	1	0	0	0	5	331	330
B 55	0	0	0	0	0	4	16	16
C 28	0	0	1	0	8	n/a	28	19
D 90	1	1	1	1	4	4	438	430
E 331	1	2	1	2	32	17	156	118
ttl	**2**	**4**	**3**	**3**	**44**	**30**	**969**	**913**
518 n	38	55	39	42	344	97	2,594	2,076
%	**5%**	**7%**	**8%**	**7%**	**13%**	**31%**	**37%**	**44%**
LISTENING	**<1 year**	**1-2 years**	**2-3 years**	**3-4 years**	**4-5+ yrs**	**exited**	**all 4th**	**non-prog.**
Dist/ ELL #								
A 14	0	2	0	0	0	6	512	510
B 55	0	0	0	2	11	5	39	26
C 28	1	0	1	0	7	n/a	40	31
D 90	5	3	3	2	11	7	679	655
E 331	2	5	3	9	72	35	320	229
ttl	**8**	**10**	**7**	**13**	**101**	**53**	**1,590**	**1,451**
518 n	38	55	39	42	344	97	2,640	2,122
%	**21%**	**18%**	**18%**	**31%**	**29%**	**55%**	**60%**	**68%**

2000-2001			Percent of ELLs in district bilingual programs meeting standard					
Notes: red n's are ELLs, blue n's are regular 4th; n/a: District C- no exit policy for K-5								
MATH	**<1 year**	**1-2 years**	**2-3 years**	**3-4 years**	**4-5+ yrs**	**exited**	**all 4th**	**non-prog.**
Dist./ ELL #								
A 17	1	0	0	0	0	3	359	358
B 87	0	0	0	0	0	7	12	12
C 36	0	0	1	2	3	n/a	11	5
D 115	0	0	3	0	4	2	475	468
E 380	0	5	1	7	28	21	183	142
ttl	**1**	**5**	**5**	**9**	**35**	**33**	**1,040**	**985**
635 n	55	63	68	78	371	113	2,632	1,997
%	**2%**	**8%**	**7%**	**12%**	**9%**	**29%**	**40%**	**49%**
READING	**<1 year**	**1-2 years**	**2-3 years**	**3-4 years**	**4-5+ yrs**	**exited**	**all 4th**	**non-prog.**
Dist./ ELL #								
A 17	2	0	0	0	1	4	510	507
B 87	0	0	0	0	1	12	21	20
C 36	1	0	1	3	6	n/a	31	20
D 115	3	0	4	6	14	6	683	656
E 380	0	5	6	12	66	42	280	191
ttl	**6**	**5**	**11**	**21**	**88**	**64**	**1,525**	**1,394**
635 n	55	63	68	78	371	113	2,628	1,993
%	**11%**	**8%**	**16%**	**27%**	**24%**	**57%**	**58%**	**70%**
WRITING	**<1 year**	**1-2 years**	**2-3 years**	**3-4 years**	**4-5+ yrs**	**exited**	**all 4th**	**non-prog.**
Dist./ ELL #								
A 17	1	0	0	0	1	4	343	341
B 87	0	0	0	0	3	12	23	20
C 36	0	0	0	4	6	n/a	28	18
D 115	2	1	4	3	3	2	493	480
E 380	0	0	5	4	41	23	163	113
ttl	**3**	**1**	**9**	**11**	**54**	**41**	**1,050**	**972**
635 n	55	63	68	78	371	113	2,566	1,931
%	**5%**	**2%**	**13%**	**14%**	**15%**	**36%**	**41%**	**50%**
LISTENING	**<1 year**	**1-2 years**	**2-3 years**	**3-4 years**	**4-5+ yrs**	**exited**	**all 4th**	**non-prog.**
Dist./ ELL #								
A 17	1	1	1	0	1	3	562	558
B 87	0	1	1	4	14	11	43	23
C 36	1	1	2	8	6	n/a	40	22
D 115	4	3	7	7	25	7	762	716
E 380	3	8	8	16	96	46	355	224
ttl	**9**	**14**	**19**	**35**	**142**	**67**	**1,762**	**1,543**
635 n	55	63	68	78	371	113	2,644	2,009
%	**16%**	**22%**	**28%**	**45%**	**38%**	**59%**	**67%**	**78%**

2001-2002			Percent of ELLs in district bilingual programs meeting standard					
Notes: red n's are ELLs, blue n's are regular 4th; n/a: District C- no exit policy for K-5								
MATH	**<1 year**	**1-2 years**	**2-3 years**	**3-4 years**	**4-5+ yrs**	**exited**	**all 4th**	**non-prog.**
Dist./ ELL #								
A 15	0	0	2	0	0	4	372	370
B 70	0	0	1	0	2	9	20	17
C 25	0	0	0	0	1	n/a	26	25
D 116	1	2	1	6	13	4	660	637
E 362	1	3	6	12	36	35	256	198
ttl	**2**	**5**	**10**	**18**	**52**	**52**	**1,334**	**1,247**
588 n	47	58	71	84	328	116	2,710	2,122
%	**4%**	**9%**	**14%**	**21%**	**16%**	**45%**	**49%**	**59%**
READING	**<1 year**	**1-2 years**	**2-3 years**	**3-4 years**	**4-5+ yrs**	**exited**	**all 4th**	**non-prog.**
Dist./ ELL #								
A 15	0	1	3	0	0	5	469	465
B 70	0	0	0	0	5	14	29	24
C 25	0	0	0	0	3	n/a	28	25
D 116	1	2	3	6	20	5	765	730
E 362	1	4	8	18	56	52	328	241
ttl	**3**	**9**	**14**	**24**	**84**	**76**	**1,619**	**1,485**
588 n	47	58	71	84	328	116	2,702	2,114
%	**6%**	**16%**	**20%**	**29%**	**26%**	**66%**	**60%**	**70%**
WRITING	**<1 year**	**1-2 years**	**2-3 years**	**3-4 years**	**4-5+ yrs**	**exited**	**all 4th**	**non-prog.**
Dist./ ELL #								
A 15	0	0	2	0	0	5	356	354
B 70	0	0	0	0	1	10	19	18
C 25	0	0	0	0	5	n/a	33	28
D 116	2	3	0	5	12	4	592	570
E 362	2	2	3	12	36	39	243	188
ttl	**4**	**5**	**5**	**17**	**54**	**58**	**1,243**	**1,158**
588 n	47	58	71	84	328	116	2,667	2,079
%	**9%**	**9%**	**7%**	**20%**	**16%**	**50%**	**47%**	**56%**
LISTENING	**<1 year**	**1-2 years**	**2-3 years**	**3-4 years**	**4-5+ yrs**	**exited**	**all 4th**	**non-prog.**
Dist./ ELL #								
A 15	1	1	5	0	2	2	478	469
B 70	0	0	1	1	11	13	37	24
C 25	0	0	0	0	4	n/a	32	28
D 116	1	2	2	9	20	5	714	680
E 362	2	4	13	13	71	49	357	254
ttl	**4**	**7**	**21**	**23**	**108**	**69**	**1,618**	**1,455**
589 n	47	58	71	84	328	116	2,731	2,143
%	**9%**	**12%**	**30%**	**27%**	**33%**	**59%**	**59%**	**68%**

All three years combined			Percent of ELLs in district bilingual programs meeting standard					
Notes: red n's are ELLs, blue n's are regular 4th; n/a: District C- no exit policy for K-5								
MATH	**<1 year**	**1-2 years**	**2-3 years**	**3-4 years**	**4-5+ yrs**	**exited**	**all 4th**	**non-prog.**
Dist./ ELL #								
A 46	1	1	2	0	0	10	1052	1048
B 212	0	0	1	0	2	18	47	44
C 89	0	0	2	2	8	n/a	48	36
D 321	3	3	5	6	19	6	1,595	1,559
E 1,073	1	10	8	21	84	75	616	492
ttl	**5**	**14**	**18**	**29**	**113**	**109**	**3,358**	**3,179**
1,741n	140	176	178	204	1,043	326	7,971	6,230
%	**4%**	**8%**	**10%**	**14%**	**11%**	**33%**	**42%**	**51%**
READING	**<1 year**	**1-2 years**	**2-3 years**	**3-4 years**	**4-5+ yrs**	**exited**	**all 4th**	**non-prog.**
Dist./ ELL #								
A 46	2	3	3	0	1	15	1,474	1,465
B 212	0	0	0	1	9	32	79	69
C 89	1	0	2	3	16	n/a	90	68
D 321	6	3	7	14	42	17	2,156	2,081
E 1,073	1	9	16	37	179	134	890	648
ttl	**10**	**15**	**28**	**55**	**247**	**198**	**4,689**	**4,331**
1,741n	140	176	178	204	1,043	326	7,953	6,212
%	**7%**	**9%**	**16%**	**27%**	**24%**	**61%**	**59%**	**70%**
WRITING	**<1 year**	**1-2 years**	**2-3 years**	**3-4 years**	**4-5+ yrs**	**exited**	**all 4th**	**non-prog.**
Dist./ ELL #								
A 46	1	1	2	0	1	14	1,030	1,025
B 212	0	0	0	0	4	26	58	54
C 89	0	0	1	4	19	n/a	89	65
D 321	5	5	5	9	19	10	1,523	1,480
E 1,073	3	4	9	18	109	79	562	419
ttl	**9**	**10**	**17**	**31**	**152**	**129**	**3,262**	**3,043**
1,741n	140	176	178	204	1,043	326	7,827	6,086
%	**6%**	**6%**	**10%**	**15%**	**15%**	**40%**	**42%**	**50%**
LISTENING	**<1 year**	**1-2 years**	**2-3 years**	**3-4 years**	**4-5+ yrs**	**exited**	**all 4th**	**non-prog.**
Dist./ ELL #								
A 46	2	4	6	0	3	11	1,552	1,537
B 212	0	1	2	7	36	29	119	73
C 89	2	1	3	8	17	n/a	112	81
D 321	10	8	12	18	56	19	2,155	2,051
E 1,073	7	17	24	38	239	130	1,032	707
ttl	**21**	**31**	**47**	**71**	**351**	**189**	**4,970**	**4,449**
1,741n	140	176	178	204	1,043	326	8,015	6,274
%	**15%**	**18%**	**26%**	**35%**	**34%**	**58%**	**62%**	**71%**

Questionnaire

District _____ **Name**_____

1. Now that you have studied the data, what do you think about them?

2. What do you think would be the most equitable way to include ELLs in the state testing system?

3. What is your biggest challenge in regards to ELLs and the 4th grade WASL?

Two further questions can be answered privately, by phone or by mail:

4. Is there anything you want to say additionally about these data?

5. Is there anything you want to say additionally about the state testing system?

Claudia Olmstead (509) 943-2681 home (509) 547-8441 ext. 246
1401 Cottonwood Dr. Richland WA 99352 *colmstead_1@charter.net*

Appendix D

List of interviews and interviewees

Date of Interview	District-People Codes:	
3-08-04	A-1	Superintendent
	A-2	Elementary Education Director
3-08-04 (written only)	A-3	State & Federal Programs Director
2-25-04	B-1	Superintendent
	B-2	Asst. Supt., Director of Curriculum & Special Programs
2-26-04	C-1	Superintendent (interviewed individually)
3-04-04	C-2	Literacy & Assessment Director (interviewed individually)
3-05-04	D-1	Superintendent (interviewed individually)
3-05-04	D-2	Migrant/ Bilingual Program Director (interviewed individually)
3-12-04	D-3	Elementary Education Director (interviewed individually)
3-10-04	E-1	Assistant Superintendent
	E-2	Director of Special Programs

Appendix E

Interview transcripts

(Note: some print has been bolded for emphasis of key points)

Person	Question	Response
A-1	1. Now that you have studied the data, what do you think about them?	Since I've looked at them J. and you haven't, would you like me to start?
A-2		I would certainly prefer for you to do that.
A-1		I found it interesting, but I didn't find it surprising. Although I wouldn't expected this well from kids that stayed in the program. *(At this point we stopped. He didn't understand the bar graphs, that the last three columns were students* ***outside*** *of the program. So I had to explain it to him and to A-2, who hadn't read any of it.)* Actually, when I looked through it I didn't catch that. What would happen if you had them exit earlier? Now the question is, should we be exiting them earlier? I mean, if you look at – look at the math here. If I were a parent, and I saw that kind of dramatic improvement, I'd be wanting my kid to exit. *(I explained to him the exit process, how students need to perform at a certain level- previously the 35th percentile on a standardized test, but now the WLPT supposedly correlated with the ITBS- so they should have to be close to grade level* ***before*** *they can exit. He hadn't been aware of this.)*
A-1		I have a basic theory about all this, I think this is far more related to poverty than it is to language. The kids we get from other countries who aren't poor, like kids from France and Germany, Bosnia, they're strong in their own language. Ruby Payne, she talks about all of this. The black kids who don't speak academic English, I see similarities…
A-2		No literature in the home

A-1		I think we'll see the second and third generation do better, though. Like the Irish and other groups. Although I think the people from Mexico are inhibited by their machismo culture. They don't seem to value education or a high school degree. In fact, if it's a girl we are talking about, they actually **discourage** her from getting a degree, because she's not going to need one in their culture. Here's another thing, J., what do you think about this? I think if we moved to a foreign country with our own kids, I think they'd pick up the language just like this (*he snaps his fingers*), because they know how to learn and they are good with language and reading and writing. They wouldn't need a special program for more than a year at most.
A-2		I am struck by the difference between reading and writing. We teach them so closely together. I would expect the reading more to mirror the writing.
A-2	2. What do you think would be a more equitable way to include ELLs in the state testing system?	We need to be clear about what we want to measure. If it's clear mathematical concepts, we need different kinds of questions, a different kind of ELL hat. We need a different kind that shows incremental growth.
A-1		If we could test in their own language, we could get to what they really understand. In fourth grade, they should be understanding in their own language. But we have, what? 26 languages here? **What we really need is the opportunity to create flexibility.**
		What people are clamoring about is that they don't want to look bad. But we do need the data in some form. Having them take the test isn't bad, other than the first year probably. It's what happens with it.

A-2	3. What is your biggest challenge in regards to ELLs and the fourth grade WASL?	Which leads to your next question. It's just such a public spectacle. So threatening – **the school's going to be taken over.** Well, they're not going to do any better. And what does WASL really tell us? The biggest challenge is this: **false assumptions.** And another thing: I have never heard any clear direction about what we can do to help ELLs in the classroom perform on WASL. Teachers get all kinds of helpful suggestions and directives for mainstream kids. But if we could hear, these are 3 strategies you can do with your ELLs to help them with WASL readiness. Have you ever seen anything like that from the state?
A-1	5. Do you have any further comments about the state testing system?	What our goal is that every student can plug into the computer whether they are here or in Seattle or wherever. We only have to be responsible for the kids who stay in our district for over a couple years or more. The state could input kids by number and follow them year by year: it's real clean, and it takes away excuses. If they move around a lot, we shouldn't be penalized. Track the kids, see how they progress. But don't hold us responsible if they don't spend much time with us.We have found a lot of problems with the WASL. The fourth grade WASL does not predict the seventh grade WASL, for instance. It's a cumbersome system. It's not timely. What we'd like to see, and what we are asking the state for, is to keep the two useful parts of the WASL: problem solving and writing. The rest can be done with the MAP test. It's immediate, it's timely. It informs teachers about the range in their classroom and the level of each kid. No one wants all these extra baby WASLs that have become big WASLs for 3rd, 5th, 6th and so on. These were forced on us.

A-2		This is the power of data. This is how we can really help kids. We can see where each kid is, and this is how we get to quality differentiation of curriculum.
A-1		We're building a strong writing program here. We have good math probes set up.
A-3	1. 1.Now that you have studied the data, what do you think about them?	I don't see any huge surprises. The other districts, due to the high percentage of Spanish language speakers, are able to offer programs that we are not able to offer due to low student numbers and a wider variety of languages enter the district. Math is not offered in our ESL program. A variety of curriculums were offered at this time.
A-3	2. What do you think would be the most equitable way to include ELLs in the state testing system?	I know it is difficult on the ELL students, the fourth grade Teachers and on the district when 100 percent of ELL students are assessed on the WASL. It is important for the ELL student to participate in the assessment for the following reasons: - provides the district to **set the baseline for growth of the student** (in the past, we were never really sure if the student made growth or not because they may have been exempted from assessment. Growth can be more actually measured over time. - Allows the classroom teacher some ownership in providing accommodations and modifications for the 4th grade student to assist the student and future students enrolled in the classroom in passing the WASL in the future. - Allows the ELL student to participate like other students in the classroom on the assessment.

A-3	3. What is your biggest challenge in regards to ELLs and the 4th grade WASL?	The biggest challenge is the research that indicates it takes 7-10 years for ELL students to acquire English (in current models of instruction) to meet standards. The challenge is finding ways to use the WASL as described above without causing the district to fail in meeting AYP with that subgroup. That needs to be amended in some way. A larger challenge looms with the Certificate of Mastery at 10th grade. Even with multiple retakes, there is an added expense for tutoring to pass the assessment and the cost of the re-takes. **Another issue is the sheer number of assessments administered to ELL students** to meet Federal Title III and NCLB requirements. They are administered the LAS (oral proficiency) and WA Language Proficiency Test (WLPT) annually. Thus, 4th grade students are assessed with four major tests that school year. That takes up a great deal of instructional time for these students.
B-1	1. Now that you have studied the data, what do you think about them?	(he hadn't had time to look at the data report yet, so he asked if we could look them over together first…he started at the bar graphs, and he asked several questions)
B-2		First and second year students looked a lot alike, and fourth and fifth year students looked a lot alike.
B-2		There was one comment in your findings that was **so** inflammatory. If it got in the hands of the wrong person it could be misconstrued.
Me:		What was that?
B-2		It was the one about how only when the students exit the bilingual program do their scores rise above the students in the program. People could take that like, "Get these kids out of the bilingual program and they'll start to achieve."

Me:		Thanks for that feedback. Help me think of a better way to say that.
B-2		Maybe mention goals of the program. Something like, "when students have met the goals of the program, they are then ready…"
B-1		So really, you aren't measuring success of a group of students from year to year. These are different groups of students.
B-2		No, you're measuring apples and oranges.
B-1	2. What do you think would be the most equitable way to include ELLs in the state testing system?	I think you have the answer right here. **You get them to a readiness level, a transitional level, first**. But not an exemption. There needs to be some kind of assessment.
Me:		How do you feel about the WLPT?
B-2		We use it here, and it's been useful. **I think if we are going to be held accountable, then we need to hold them in the program after they qualify for exiting.**
B-2	3. What is your biggest challenge in regards to ELLs and the 4th grade WASL?	**There are so many. How can we pick one?** There are kids barely coming in that we have to test right away. There are kids with no English and no former schooling. There are homes with no literacy. There are homes of extreme poverty. There are teachers with too much to do already. But I guess the biggest challenge is the SPED and the LEP kids. It should not be automatic for them to take it. It's ridiculous. They're just not ready so much of the time.
B-1	4. Is there anything you want to say additionally about these data?	**Longitudinal data is much more meaningful.** We've studied our own kids; in fact, we have a good little study where we followed our first group of students who took the WASL in 4th grade, 100 of them. We captured their scores again in 7th grade and 10th grade, although we were down to about 68 by that time. But that kind of study shows us a lot more.

B-2		We've been able to show tremendous success that way. With our kids here, we have to aim for two years of growth every year. We're getting that, with our 5-week summer school, breakfast reading club. We can show one year of growth in reading comprehension in five weeks of summer school, but parents complain that their kids don't have any fun. We have Reading First this year, and in half a year kids are showing a year's growth as well.
C-1	1. Now that you have studied the data, what do you think about them?	Well, I would have thought by 4 or 5 years there would have been more improvement. Then of course I always want to go to the data.
Me		How so?
C-1		How did we compare with the other districts? Like I look here in reading, at the kids who have been in program 4-5 years, in the first year. (So we did some math, and discovered that the students in his district percentage-wise did better than the other districts.) Say we take the n's of everyone and put them in percentages, maybe the ELLs wouldn't be too different from the whole population here. (So I told him how that would be on the original data charts C-2 had done for me.) My theory is that in 1 or 2 years, chances are almost nil. And here, they're less than 10%. Even at 4-5 years, they're barely a quarter. That's telling. Another thing, I'd like to see stats on free and reduced lunch- how it fits.

C-1	2. What do you think would be the most equitable way to include ELLs in the state testing system?	**They should not be tested until they're proficient.** It's unfair to test them in the language they are not proficient in. Until we can guarantee that they are proficient in the language, they should not have to take it. **It's not fair, and it's not achievable.**
Me		Would you assess their proficiency with the WLPT?
C-1		Probably not. Some standardized test, but that **WLPT takes so much time out of the year, out of students' time, out of teachers' time.** Something quicker. Something computerized would be better maybe. Our bilingual teacher just told me that a classroom teacher told her, I will not send my kids out again for (WLPT) testing. I spend a lot of time preparing. You want to pull out my kids, you better bring the superintendent with you.
C-1	3. What is your biggest challenge in regards to ELLs and the 4th grade WASL?	It would be the lack of experience. The typical child is migrant, they're new to the country, they're **3 years behind academically**. They are not prepared with the experiences they need.
C-1	4. Is there anything you want to say additionally about these data?	I think these data need to be shared with the people who make the policies. They need to see this. I haven't seen anything that breaks it up this way, but this study is objective, it shows small, large, across a large area, here it is. This is what happens. Not in one, not in two years…it can't happen.
Me		What do you think about the reading scores showing more of an upward growth?

C-1		That's because we've been focusing on literacy for the last few years. It shows. Now we've been focusing on writing, and that's showing more too. I guess one more thing – Preparation is the big thing really. Like attendance. How many days did they attend or not attend? Because **you'll often find they did not attend an academic year.** I'd like to do a random sampling and see how many days are averaged among our students. And the home life. What the home life is like really influences preparation for school.
C-1		What are you going to do with this? I think you ought to be showing it to everybody you can. People need to be seeing this.
C-2	1. Now that you have studied the data, what do you think about them?	What it really means to me is that we shouldn't test kids on the WASL until they exit from the program. Take what we already know and put it right in front of our face. I mean, you can't argue with the data. This here is quite dramatic.
Me		What do you make of the bar graphs?
C-2		It's the **learning process, exactly as we understand it. One year of learning, a following year of maintaining.** You see a year of stretching, you see a year of letting it sink in. I mean, we know this from brain research. I think you did a good job of putting this together. Your data are very illustrative of your thesis.
C-2	2. What do you think would be the most equitable way to include ELLs in the state testing system?	On the basis of your research, I would say they need to be out of the bilingual program first. It would be interesting to see what would happen 2 years after they exit, though. How would their progress continue?
Me		Collier and Thomas' research shows that if they only make one year's progress a year, they will fall further and further behind each year.

C-2		Yeah. That makes sense. That's not a good thing.
Me		What would you use to assess progress of ELLs while they were still in the bilingual program?
C-2		Well, not the WLPT. It stinks. It's not a format they're used to, it assesses a different kind of thinking than what we work on in class. I think an oral test at first is better. The LAS-O is probably best, but it shouldn't be the only point of data. Really classroom-based assessment with a clear target, goal setting process...that's how you could show **progress.**
C-2	3. What is your biggest challenge in regards to ELLs and the 4th grade WASL?	The first challenge is, getting people who can work with the students in their home language at an academic language level. It's just a bandaid to get people who can speak their language but not really read or write in it. **Another challenge is that after the test, their self esteem goes way down. It's like if someone gave us a test in Chinese and said, here, go take this. And the test takes such a big chunk of time. It causes just a big down shift in the learning momentum..** But the answer, the biggest challenge, is that we need an **appropriate** measure to show growth. WASL is not the answer. **What are you trying to assess there? Content or language ability? Seems to me to be language ability, because you can't even get the kids to the content. The issue is, what's appropriate.**
C-2	4. Is there anything you want to say additionally about these data?	This needs to be published. It's important information, and it needs to be assertively distributed. By that I mean you need to go to conferences, put abstracts in teaching journals and administrative journals so they can contact you if they're interested, send this report to our Secretary of Education.

C-2	5. Is there anything you want to say additionally about the state testing system?	That this was supposed to be about classroom data. This was never the point. The point was to ratchet up the instruction, not to make everything completely WASL-driven.
D-1	1. Now that you have studied the data, what do you think about them?	Well, there are some concerns with the data. The big issue is, how many kids were at level 1 and so on? I mean, we have some kids exiting, and they are doing very well. What level did they come in at? But I'm a little concerned. We have some kids coming in, and after 3 and 4 years in the program, they appear to have dropped off. **What are the percents of kids who came in as monolinguals? What percent of these are kids who aren't very fluent in either language? I mean, you would have a huge study here. These are all your variables, though.** I guess the thing in general that interests me the most is, what happens between years 2-3 and 4-5? Because there appears to be some falling down here. What is the rationale around that? Because some kids did better in years 2-3 that they did in years 4-5. The other interesting thing is how well the kids who exited did. Those kids may be your more precocious learners. Do they have more propensity for learning language?
D-1	2. What do you think would be the most equitable way to include ELLs in the state testing system?	We need to buy the kids some time. The question is how much time are you going to give them. Because clearly, **it's a political issue, not an educational issue. No matter how you cut that, politics is going to rule. If we could get 3 years, I think that's the best we can get. It's not necessarily an equal chance, but it's a better chance.** **Most of these decisions are not made on research of child development, they're made by politicians.**

D-1	3. What is your biggest challenge regarding ELLs and the 4th grade WASL?	How quickly we can get the kids to learn English. We're doing a lot of work in reading. Reading in Spanish. You've got to remember reading is not the issue. When we get the kids reading in Spanish, or their primary language, it's also the transfer of skills. The fourth grade WASL is really a fifth grade test. Those kids have to know fifth grade vocabulary and material. And that's even more true in Math. The math is end of fifth grade level. But maybe a bigger challenge is that no one knows how to deal with the students who are not proficient in either language. Not proficient in vocabulary or structure. And that can be a native English speaker, as well, who is not proficient in English. What can you do with them, what is most efficient, how can you get parents to understand that language development is so crucial? I guess what we've decided, is they're predominantly English, we'll work with them in English. If they're predominant in Spanish, we'll work with them in Spanish. But I don't know if that's a clear answer. All of it is challenging.
D-1	4 - 5. Is there anything else you want to say about the data or the state testing system?	There are weaknesses with the state assessment system. It's a combination of educational and political decisions. There is a fairness issue with the English language issue and these kids. But the other side of it is that it does force us to look at **instructional technology**. We're doing a better job here because of it, better than we've ever done before.

D-1		I wish we knew more about how we could help kids learn a second language quicker. I know there's a lot of argument about that. You can't get researchers to agree. But **I think there'll be a breakthrough and we will learn ways to do it faster.**
D-2	1. Now that you have studied the data, what do you think about them?	There really weren't any surprises. Just looking at the regular kids who are taking the test, they're having trouble. And then you take the kids whose first language is not English. **They need more time, and that's the bottom line.** Ultimately, it's the tenth grade when we have to really worry…I'm not for **not** testing these kids. I'm having a lot of different feelings. A lot of these kids, they might have been at the top of their class in their home country. But they can't do what's required on the WASL. I'm not opposed to testing them, but I don't think it's fair to them. I go into their classrooms, I've seen it too many times. The look on their face, like "I'm stupid". They're not. I saw in the paper this morning, they're (the Washington legislature) proposing some changes. I'm anxious to see what they are. The secretary of education announced the new changes- but one year's not enough. Like I said, I'm not opposed to the testing. But it's so hard on these kids.
Me		Did it surprise you about the kids that exited the program doing so much better on the WASL?
D-2		I'd be interested to see the names of those students. I'm wondering if they came from a good education in their home country. Because that's the trend that we find, the stronger they are in their home language, the better they do. But some of these kids, they're so bright…when they come in, they're in ESL for a year and then, they don't need ESL any more.

D-2		It should be getting a little better. We're doing better every year. We're teaching them reading in Spanish now, and in third grade, we're using a new Transitions program. And they transition to English reading. This is our first year. We use a Levels Reading test. We just finished our mid-year testing, and it looked good. We're seeing big results. Then next year, when we take the WASL, we'll see. **It's been good, this reform's been good, forcing everyone to look not just at ESL kids. I've been here since '81, and I've watched, you know, oh those kids over there.**
D-2	2. What do you think would be the most equitable way to include ELLs in the state testing system?	Well, like I said, I'm not opposed to the testing, but I'd like to see them not included in the district scores. For how long, I don't know. I think it's kind of unfair to the district to include these kids, because of, you know, the language. I know the state can't come up with a test in Spanish or a test in Russian, because we have too many languages. And often the kids don't have the academic language in that first language anyway. **So it has to be in English, we know that. But it still doesn't seem fair to penalize the district or the ELL kids because they haven't had the experience yet to pass the test. We need to give these kids extra time.**
D-2	3. What is your biggest challenge in regards to ELLs and the 4th grade WASL?	The pressure for them to perform like native English speakers, and it's all AYP. Principals are all up tight, and the kids are nervous. We've talked to the parents, and the parents are not opposed. But some of them say, "but they don't even speak English!"

D-2		But I think the biggest challenge is trying to give these kids more time. Dealing with AYP on the one hand and trying to help the kids relax even if they just got here from Mexico. Some of the teachers will have them just go ahead and write in Spanish so they can feel good, like they were able to do something.
D-2	4. Is there anything else you'd like to say about the data?	No I think I've already covered that.
D-2	5. How about the state testing system?	We need to look at something different. Maybe not at the same standard, but maybe here's where we expect them to be at one year, and at two years, and eventually meeting up with the WASL. We need some sort of standards so that teachers can see where they're going. I'm excited about what I'm seeing. We're doing new things and seeing some results. ELL kids are at the bottom, it's discouraging. But at least if we had a scale, and if we could see progress and a goal to reach to, then I think teachers wouldn't feel beat up, kids wouldn't feel beat up, parents wouldn't feel beat up.
D-3	1. Now that you have studied the data, what do you think about them?	I thought it was really interesting. In fact M. (D2) just gave me the other summary that you've done-
Me		The breakdown into 1, 2, 3, 4, years? (I had given them the more detailed district report on request than the data summary report)
D-3		What's interesting is it makes you wonder if we've plateaued a little bit. The numbers have increased, but the percentages haven't increased over the last couple years. Can you see that? When you look at the percentages, we went up significantly from 99-2000 to 2000-2001, but then we're basically kind of flat in terms of percentages.

D-3		Then the last couple years in math we went flat too, I mean the numbers increased, but we went flat in terms of percentages. I just think it's interesting, I don't know what more to make of it right now. We also did a little bit of a study of our own a couple years ago when we did our levels testing [reading], we only pulled out kids who'd been in our bilingual programs, and to be honest it was a little discouraging. Although there's some question about the validity of what we did. And in that one, a couple of schools did really well long-term. Because what we did was similar to what your study is, you know, what's the long term effect of bilingual ed, is that what you're looking at? No, you're looking at WASL…
Me		No, I'm looking at time. How much time do they need? Is all the previous research right. It tells us 7, maybe 10 years.
D-3		We didn't find, when we pulled out levels data, we found that there was one school that was very effective.
Me		What was your really effective school?
D-3		Ridgeview [a school in a very affluent neighborhood]. But what we don't know, the piece we didn't look at, was where did everybody start? Does it make a difference if we have some real low kids starting? And these kids are just higher and naturally grow more? I don't know if that makes a difference.
Me		I think it does. Research says that it does. The most important variable is supposed to be where they start from and if they have literacy already in their first language. It's kind of the whole package, you know: what's the socioeconomic status of their family, what's the educational background of their parents, do they have literacy in their first language or if they don't are their parents trying to build it at home while they're learning English at school.

Me		And then you have the Ridgeview neighborhood. And then you have another variable, which is they probably have less ELLs than at the other schools. Research shows that the more ELLs you have at a school, the lower your scores are going to be, which is also tied into SES. The state has done some research on this.
D-3		Then there's the issue of English. Do we believe the kids have to learn English to be successful? I'm just curious- what's the variable that makes Dual Language successful? There's very few pure models out there. I talked to a guy from Grandview where they have a real successful program. They do a model that's 80% English and 20% Spanish. And then in Pasco, they're talking about doing the exact opposite- 80 Spanish, 20 English. I'm real curious about that. And then the other side of that is the reading research says that kids need to be reading by the time they leave second grade. So if we're messing around with oral language stuff at kindergarten and first grade, how's that going to be reflected?
Me		Any other gleanings?
D-3		Actually, you go down the list, this is percent meeting standard, and it builds pretty quickly for the first couple years. This does seem to plateau a bit too. What frustrates me is, I think there's more that we don't know about second language kids than what we do know. For example, for us, the kids that come from the Eastern European bloc countries attend one school for the most part, and they do pretty well after they've been here for awhile and their scores tend to continue to grow. But the question is, what's the variable?

D-3		Most of those kids have been to school, most of those kids are refugees, and I think they come from a little more affluent families in their home countries. Whereas the kids who come from Mexico are basically coming from poverty. And I think their parents think they've got it pretty good up here even though by our standards they're having a tough time. But they've got carpet on their floors instead of dirt and they think, we've got it pretty good. There's just those variables, how do you measure them?
		(We just looked at the data for awhile and talked about the study design and the different districts.)
D-3		Did you do anything here that compares instructional models or service models at all?
Me		No I didn't.
D-3		What I'm just really really curious about, because it's what we've been working on, is the issue of instruction. Because, I hate to say it, but I think the sad truth is, especially in 99-2000, at least in our program, I'm not convinced these guys were getting the best instruction. And one of the things that M. and C. have really done, we have really tightened up our curriculum. And also, the other thing that we struggled with, back especially in these years, is we didn't have the materials to support the instruction. And we bought, this year, we use Open Court in most of our other schools, and we bought, it was supposed to be a parallel program in Spanish to Open Court but it stunk. It was terrible. Our teachers complained to Open Court, and the next year, they redid the program.

D-3		Well, this year and now it's a much better program. It's got similar supplemental materials, similar intervention materials to what Open Court has, so part of my question now is, so this percent, **I'm thinking, is now we have said, 'here's your language program, here's your reading program, here's when and how we want you to transition kids.'** C. has even found a program or a set of teacher materials anyway that says here's the problems that kids are going to have as they transition from language to language. And here's how you can deal with it. I think that, especially our reading scores, are especially going to jump. We've got some great teachers for those kids. I heard Ed Camaren (?) from the University of Oregon talk, and this thing has always stuck with me: **"A teacher can't be the conductor and composer of the curriculum."** The teacher's job is to conduct the materials so they meet the needs of the kids. The district and the administrator's job is to help the teachers find the very best materials and the art of teaching is to use those materials so they meet the needs of the kids. But he says we can't expect teachers to create all this stuff, because they'll go nuts. But back here (1999-2000), if you were a bilingual teacher, you had to, because it wasn't available. I'm more excited about what we have now to work with. Matter of fact, our mid-year levels test scores on our second language kids, we had more kids meet standard than we've ever had. So we're thinking this spring, although it may not be real high percentages of kids there, but it's going to be better. Significantly better.

D-3	3. So, then would you say your biggest challenge with ELLs has been getting the right materials for them?	Yeah, I think one of the problems with second language instruction has always been that there haven't been enough materials. My bias is that teachers need that. Just like with a carpenter, you need the very best saw. It's still how you use the saw. And then the other thing is, you systematize the program you're dealing kids, because when Teacher A is having to make up the curriculum, as much as they try, it's not going to flow into Teacher B's classroom next year, because their knowledge, skills and experience are going to be different. So when we systematize it, the kids' experiences become very similar. And especially for these kids, who have to learn twice as much. We have to do everything we can to support them.
D-3	2. What do you think would be a more equitable way to include ELLs in the state testing system?	I think we need to keep giving them the WLPT, because that sends a message to everybody: that the goal is to get kids into English. **I think we need clear standards based on the WLPT about when a kid is ready to take the WASL in English. And it's got to come off a test like the WLPT, because otherwise districts are going to interpret that differently**. We're going to be compared on the basis of our WASL tests, no matter what anybody says, we're going to be compared. And so what's embedded in that is, there are some kids that need to take the test in another language. We give the DRA in Spanish, and we have kids who are reading up a storm in Spanish. So this test, these scores reflect who meets the standard on the DRA. When we first started doing this, it moved from 35% to 70% meeting standard [he's showing me a printout of first graders]. But when they get to third grade, and they have to take the levels test in English, this is going to drop down to 30%.

D-3	Is there anything else you want to say about the data, the WASL?	The only other thing is, it'll be nice when we can get with our second language population where we are like with reading, when somebody comes out and says, you know, we've taken a broad look at the studies that are out there, here's what we think, Here's the elements of instruction these kids need. Because right now, it's pretty mixed, and it's really hard to say it's a clear path. It really depends on who you listen to.
E-1	1. Now that you have studied the data, what do you think about them?	I guess when I look at the charts, the bar charts, it reflects a lot of the data that L. (E-2) has given to us over the last couple years or over the last few years, there have been some dramatic increases. And also the disparity between the program and non-program kids. And especially if you look at some kids that have been with us longer, we see that kind of growth. And we feel like we've taken some steps, but we have some new steps we're working on. L. has spent a lot more time with this than I.
E-2		Well, as I look through the students that we sorted over the years, it always made a difference in when those kids came to us and the background that they have when they come to us. So for instance if they're a fourth grader and they've only been here a year, and they've been to school before and they're fluent readers, they are doing sometimes surprisingly well for the limited amount of time that they've been in English. And then again we might have a child who has come to us in the recent years and who has never been to school and we know it's going to take a lot longer for them. That kind of information wasn't surprising at all, but looking at each individual student when I was sorting the data...

E-2		The other thing that I noticed was that those kids who looked like they may have been with us for a long time had breaks in their education. So they may have left us every year for a couple months, and that made a difference as well. So it's not easy just to say, they've been here 4 years or they've been here 2 years they should have this amount of growth. It all depends on their background and the consistency they've had in their education as well. A lot of our kids, too, it depends on their home environment. Many of our kids who qualify for our Spectrum, our sort of gifted program here, in the bilingual piece, look a lot like a middle class white student in their environment. They have two parents, they've read with them at home in their home language, so they're much more advanced in their literacy and vocabulary skills. Where a child who comes from a parent who isn't a reader and a writer or a limited reader and writer, and who doesn't experience the same kind of environment, it takes much longer for them to progress in literacy and math and everything the WASL expects. That kind of influences and changes what happens when you look at individual students, so **it's more than just looking at the child, but looking at their background and their parents.** And their parents aren't always truthful about their literacy skills.
E-1		The common denominator there is poverty vs. non-poverty and stability in the family and educational background. As L. mentioned, our Spectrum program is our gifted program, and we have a bilingual side of it, and then our English side of it so that kids are identified and put into it. So for example, that bilingual kid looks different from that bilingual kid down the hallway, literally.
Me		When did you guys start that?

E-2		He started it.
E-1		Four years ago, maybe?
E-2		I think this is our fifth year.
Me		Now I know you've been following your exited kids to see how they've been doing…
E-2		It's our dipstick approach to see that they have sustaining success, that it isn't just through the program, that when they exit and leave into the mainstream that they continue their academic success. And then I feel that the program's been successful, that we haven't let them go too soon, that they've received the kind of support that they've needed. I'm finding that the majority of the kids are successful, and that's why we do that. We don't look at them individually, they do that out at the buildings. But at this level, I disaggregate their test scores to see if they're following the flow of the district, and they usually exceed the district average or at least right there with the district average.
Me		So in view of the new provision to give them 2 more years, to include their scores with the ELL subgroup, do you think that will help your subgroup score?
E-1		Who knows? It's hard to tell. **The bottom line is, we're giving a test in English to kids who are learning English, so at what point in time is that equitable?** Is two years the magic number? **It's problematic no matter what time frame you put around it.** Because we know, our kids that are speaking Spanish, you give them a test in Spanish, we get a very high rate of kids scoring at a high level. So, you turn around and require them to test in English, prior to them being ready for a test in technical English, and their scores go down. And so there's a quandary until somebody realizes that, yeah so it may be helpful, but what does that really mean.

E-1		I sit on the state AYP and Special Ed committee and we're trying to pull both Special Ed and LEP out, because those are programs. Those kids have been identified for special assistance. So how do you test them equitably with anybody else? You can't. But it's like your data are showing, kids who have exited from the program are scoring at a fairly high rate.
E-2		The reason that they've exited is because they've met a criteria of high test scores. So we just hope they continue that.
E-2	2. What do you think would be the most equitable way to include ELLs in the state testing system?	Well I think what we're attempting to do across the state is to test kids against themselves, so, with all the different variables that we talked about at the beginning, is the students may be here for 2 or 3 years, but if they've had a break in their academic year, that makes a difference, so we can't just say so they've been in Washington schools for 3 or 4 years, they should progress like all other kids. It's very individualized. The other part is that those kids who have exited, we have never been able to disaggregate those kids away from the English-speaking population and say this is a success. Those kids who come to us and move into the mainstream, they just now get counted like everybody else, and that's not truly who they are. It takes a long time to acquire English at a high technical academic level. So they're still continuing to learn English throughout their years. But at least we can show that when they've left the program and they're no longer considered program kids that they're successful and so there have been successes at the district level. Limited English kids are a revolving door. It's hard to judge a district and what they've done for students based on the newcomers all the time. That's why it's so important to include the exited students as a disaggregated component.

Me		Now you're on some state committees too, aren't you? You're on the ELD Standards (English Language Development) committee?
E-2		And the Bilingual Committee, and the Bilingual Special Ed, Certificate of Mastery…
E-1		She's actually OSPI East, when it comes to this arena. And they're listening to her finally, after all these years.
E-1		One of the other variables that a lot of people don't recognize is, for us, look at your samples and the n's within the samples, and this is just one grade group, we have a significantly larger group of students that our district is dealing with. I mean we can call kids LEP but there are all shades within that group. And a revolving door group. So we're talking 3-4-5 thousand kids, easily, that language is one of the issues that they're working with to be successful at school. And the other sites are maybe talking 100 kids, or 50 kids, or 10 kids, so it's a whole different dynamic and challenge.
E-2		Well, one thing I know that the feds are attempting to do and we are at the state level is with the standardized assessment for limited English students, and assessing them against themselves, and how they're growing in their English language, and getting closer to meeting the standards, the ELD standards, would be of great assistance. We selected a test before we had the standards, which was backwards. Now we have the standards, we need to look at an assessment tool. But the feds have asked us to take a look at kids against themselves. So what the state with a unique identification number for every student will be able to track those kids, so we can see how many moves they make, so we can see the challenges that each student faces. So it's not excuses, but it's better understanding. Again, it's back to individual students.

Me		Do you think we will keep the WLPT?
E-2		Well, whatever we pick will continue to be called the WLPT. Every publisher is contacting us to try to get us to pilot their test. That's the last thing we need, is more testing. I've looked at some others that other states use, and I think, I have not seen an assessment out there yet…
E-1	2. Do you have any reflections on more equitable ways to test ELLs?	L. has been involved in that and reflects what we think.
E-2		I think we need to assess our kids in both languages, so we can monitor their progress in each. But then that gets to be the challenge, it seems like all we're doing to our kids is testing them. So I think there does need to be assessment in both languages. But we're unique in the state: there aren't that many districts that provide instruction in the students' primary language. So for us, that would be great, to have assessment in two, actually three, languages. [Russian is their third language.] But that would be a costly endeavor for the state, especially when so few districts in the state have primary language instruction.
E-2	3. What is your biggest challenge in regards to ELLs and the 4th grade WASL?	**Time to show proficiency**.
E-1		And your research shows that.
E-2		For the rest of our students, our challenge is literacy, and whether there's support for that at home. We have found again, that what seems to work is looking at the individual child.

E-1		The key for that lies with instructional strategies, more than there's a book or a magic program per se that you can buy off the shelf. It's teaching teachers how to be more effective, so that they are looking at each child, looking at each strategy. It's knowing the strategies and techniques, and knowing how to identify what the students need. I think in the past, in general, we have gotten caught up in buying a book or a curriculum or a program off the shelf, and haven't given teachers the tools and instructional strategies to really instruct kids and identify strengths and weaknesses. And then what kinds of materials to make that kid a better reader.
E-1	4. Is there anything further you'd like to say about the report?	It reflects, which is a positive, what we see. The growth over time. Reinforcer or reinforcement of what we've already found to be useful, a reaffirmation that somebody outside interprets the data the same way we are.
Me		I'll tell you, I really tried to find districts that were convenient, but that also really represented what there is in our state. And you definitely represent a very impacted district. And I really thank you.
E-2		I tell you why it took me so long to respond to some of these. When I went back and I thought I had lost those data that I had pulled for you, I started to go back and redo it the same way. I would find a dozen or so kids who would show an original test going back to kindergarten, and then when I went back to make sure I'd find 2 or 3 years that they weren't here. And then when they'd come back in, they'd be lower then, well they'd still be low. And I'd think, this doesn't look right, and I'd have to go back and count the years they were missing. So that was – as you know, with the number of kids that we have, it's not easy to track them all back.

E-2		There's probably three quarters of them that were a straight flow. So if they missed 1 or 2 months, I didn't catch that. But for those that had a year missing, or 2 years or 3 years, I had to go back and find them. So that was the challenge. I feel it was accurate, though!
E-1	5. Is there anything you want to say additionally about the state testing system?	**We have no problem with assessing kids' progress and being held accountable to that, but we want to make sure that the assessments are equitable and that they fit the kids that are being tested.**

Appendix F

Themes and sub-themes from interviews with district administrators

Theme	Sub-theme	District				
		A	B	C	D	E
Reactions to data report	Data reflected what they are seeing. Students are making progress.					E1
	Surprised at how poorly students in program 4th and 5th year did. **It builds pretty quickly for the first couple years** ***(progress)*****, and then it plateaus. It's pretty frustrating.** (D2) **I have some concerns…There appears to be some falling down here. Some kids did better in years 2-3 that they did in years 4-5.** (D1)	A1		C1 C2	D1 D2	E1
	I'm not convinced these guys were getting some of the best instruction. Something that we really struggled with, back in these early years, we didn't have the materials to back up the instruction.				D2	
	Concern about the long-term effect of bilingual program.				D2	

	What would happen if you had them exit earlier? Now the question is, should we be exiting them earlier? I mean, if you look at – look at the math here. If I were a parent, and I saw that kind of dramatic improvement, I'd be wanting my kid to exit. *(I explained to him the exit process, how students need to perform at a certain level- previously the 35th percentile on a standardized test, but now the WLPT supposedly correlated with the ITBS- so they should have to be close to grade level* ***before*** *they can exit. He hadn't been aware of this.)*	A1				
	Thought it could be interpreted wrongly by some: **"Get these kids out of the bilingual program and they'll start to achieve."**		B2			
	Important variables not included – amount of English students came in with, fluency in L1	A1		C1	D1, D2, D3	E2
	Very interesting how well the students who exited did	A1		C2	D1 D2	E1
	My theory is that in one or two years, chances (of meeting standard) **are almost nil. And here, they're less than 10%. Even at 4-5 years, they're barely a quarter. That's telling.**			C1		
	I think these data need to be shared with the people who make the policies. They need to see this. This study is objective, it shows small, large, across a large area, here it is. This is what happens. Not in one, not in two years…it can't happen.			C1		

	Take what we already know and put it right in front of our face. I mean, you can't argue with the data. This here is quite dramatic.			C2		
	(Bar graphs) **It's the learning process, exactly as we understand it. One year of learning, a following year of maintaining. You see a year of stretching, you see a year of letting it sink in. I mean, we know this from brain research.**			C2		
Factors affecting WASL achievement	Literacy in the home	A2	B2	C1	D1	E2
	Poverty	A1, A2	B2	C1	D2	E1 E2
	Industrialized countries education systems/ poverty issues **For us we have kids from the East European bloc countries that attend one school mostly, and they do pretty well after they've been here a while, and their scores tend to continue to grow. But the question is, what's the variable? Most of those kids have been to school, and are refugees, and I think most of them have come from a little more affluent families in their home country, whereas our kids from Mexico, most of them are basically coming from poverty.**	A1			D1 D2	
	I have a basic theory about all this, I think this is far more related to poverty than it is to language. . The kids we get from other countries who aren't poor, like kids from	A1				

	France and Germany, they're strong in their own language.					
	Home environment			C1		E2
	No former schooling		B2		D3	E2
	Different experiences: typical child is migrant, new to the country, on the average 3 years behind academically			C1		
	Language development in L1				D1, D3	
	Teachers with too much to do		B2			
	Absentee rate (what's counted as an academic year often isn't, because students have been in and out of school) Can have a negative effect on achievement (i.e. students will regress) (E2)			C1		E2
	Instructional strategies, especially for students with enough English skills and academic readiness					E1
WASL not appropriate test for ELLs.	**We need a different kind (of test) that shows incremental growth.** Need different type of standards for ELLs, so teachers can have a standard of expected yearly progress, that eventually meets up with the WASL. (D3)	A2	B1	C1 C2	D3 D2	E2
	I think what we're attempting to do across the state is to test students against themselves. So all the different variables we talked about at the beginning, the students may be here for two or three years, then if they have a break in the academic year that makes a difference and we can't just say, they've been in Washington schools 3 or 4 years, then they should progress just like other kids. It's very individualized.					E2
	ELLs and SPED shouldn't be		B2			E1

	included automatically in WASL					
	Measures language not content			C2		
	Fairness issue			C1	D1, D3	E1
More time; wait for more academic readiness		A3	B1	C1	D3	E1 E2
	Give them 3 years 1 year **It's a political issue, not an educational issue. No matter how you cut that, politics is going to rule. If we could get 3 years, I think that's the best we can get. It's not necessarily an equal chance, but it's a better chance.**	A 1			D1 D1	
	Get them to a readiness/ transitional level first, measured by an assessment. No exemptions, though.		B1			
	Transition them from the program first. (C2) **Like your data show, that kids that exit from the program are scoring at a fairly high rate.** (E1) **The reason they've exited was because they've met a criteria of really high test scores** (E2).			C2		E1, E2
	Until we can guarantee that they are proficient in the language, they should not have to take it. **It's not fair, and it's not achievable.**			C1		
	I think we need clear standards based on the WLPT about when the kid is ready to take the WASL in English. And it's got to come off a test like the WLPT, because otherwise districts				D2	

	are going to interpret it differently.					
Don't include their data in reports	Test them, but don't penalize them or rest of kids for their low scores				D3	
Language of test	Testing in L1 would be useful, but not feasible. 26 different languages in our district. (A1) Too many languages. (D3)	A1			D3	
	Kids that are stronger in Spanish or Russian need to be tested in Spanish or Russian.				D2 D3	E2
Too much time spent on testing	**Sheer number of assessments administered to ELL** students to meet Federal Title III and NCLB requirements. In District A, 4 th grade students have to take 4 major tests in a school year.	A3				
	Teachers complaining (to the superintendent) about how many tests the ELLs have to take.			C1		
	Too much time	A3		C2		
After redesignation	Hold in program after exiting. If we're accountable, then we're accountable for them when they transition out.		B2			
	We need to follow them to see if their progress continues.			C2		E2
Testing itself not bad.	Having them take the test isn't bad, other than the first year probably. It's what happens with it.	A1				
	Need a baseline, helps teachers arrange for modifications on further assessments, allows students to participate w/ peers.	A1, A3				
	It has improved our instruction				D1, D3	
	They should take the tests the other kids take.	A3			D3	
Testing can be harmful	Very hard on self-esteem			C2	D3	
WASL data not	What does WASL really tell us?	A2				

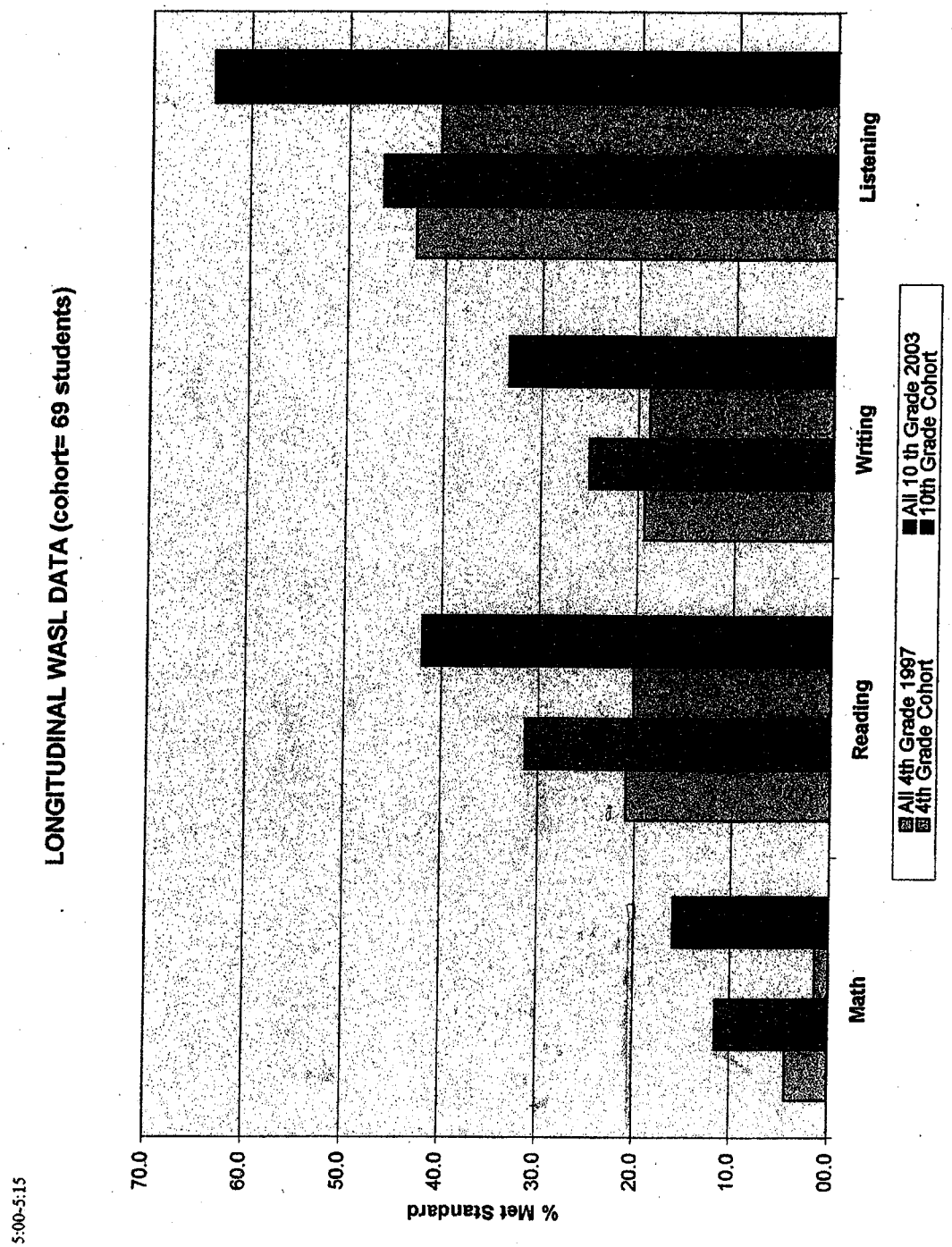
helpful	**false assumptions.**					
	Fourth grade WASL doesn't predict seventh grade WASL	A1				
	WASL doesn't give any suggestions to teachers re: how to help ELLs perform on WASL. Lots of suggestions, materials for mainstream kids, however.	A2				
	Not timely, doesn't help teachers know about the kids they have in class.	A1 A2				
WASL has problems	**It's just such a public spectacle. So threatening – the school's going to be taken over. Well, they're not going to do any better.**	A2				
	State rather than schools should be responsible. Students transferring from town to town should be tracked by state. **We only have to be responsible for the kids who stay in our district for over a couple years or more. The state could input kids by number and follow them year by year: it's real clean, and it takes away excuses. If they move around a lot, we shouldn't be penalized.**	A1				
Challenges	**No one knows how to deal with the students who are not proficient in either language. Not proficient in vocabulary or structure. And that can be a native English speaker, as well, who is not proficient in English. What can you do with them, what is most efficient, how can you get parents to understand that language development is so crucial?**				D1	
	I think there's more about				D2	

	second language learners that we don't know than what we do know.					
	I wish we knew more about how we could help kids learn a second language quicker. I know there's a lot of argument about that. You can't get researchers to agree. But I think they'll be a breakthrough and we will learn ways to do it faster.				D1	
	It would be the lack of experience. The typical child is migrant, they're new to the country, they're 3 years behind academically. They are not prepared with the experiences they need.			C1		
	AYP and more time				D3	
	Time					E1, E2
	ELLs are just a revolving door. You really can't measure them.					E2
	People who can instruct the students in their primary language and an appropriate state assessment to measure their progress			C2		
	One of the problems with second language instruction is that there haven't been enough materials. Just like a carpenter needs a good saw. It's still how you use the saw. And then you systematize the program that you're giving kids, so the kids' experience becomes very similar. Especially for these kids, that have to learn twice as much. We have to do everything we can.				D2	
Problems with study	Have to be careful with presentation. In the hands of wrong person, could be		B2			

	construed that kids are held back by the bilingual program, since their achievement upon exit takes a giant leap.					
	You're measuring apples to oranges here.		B1			
	You can't just make conclusions based on time. There are shaping factors with a high degree of influence (language development, amount of English they came in with, former schooling, absenteeism)		B2		D1	E1, E2

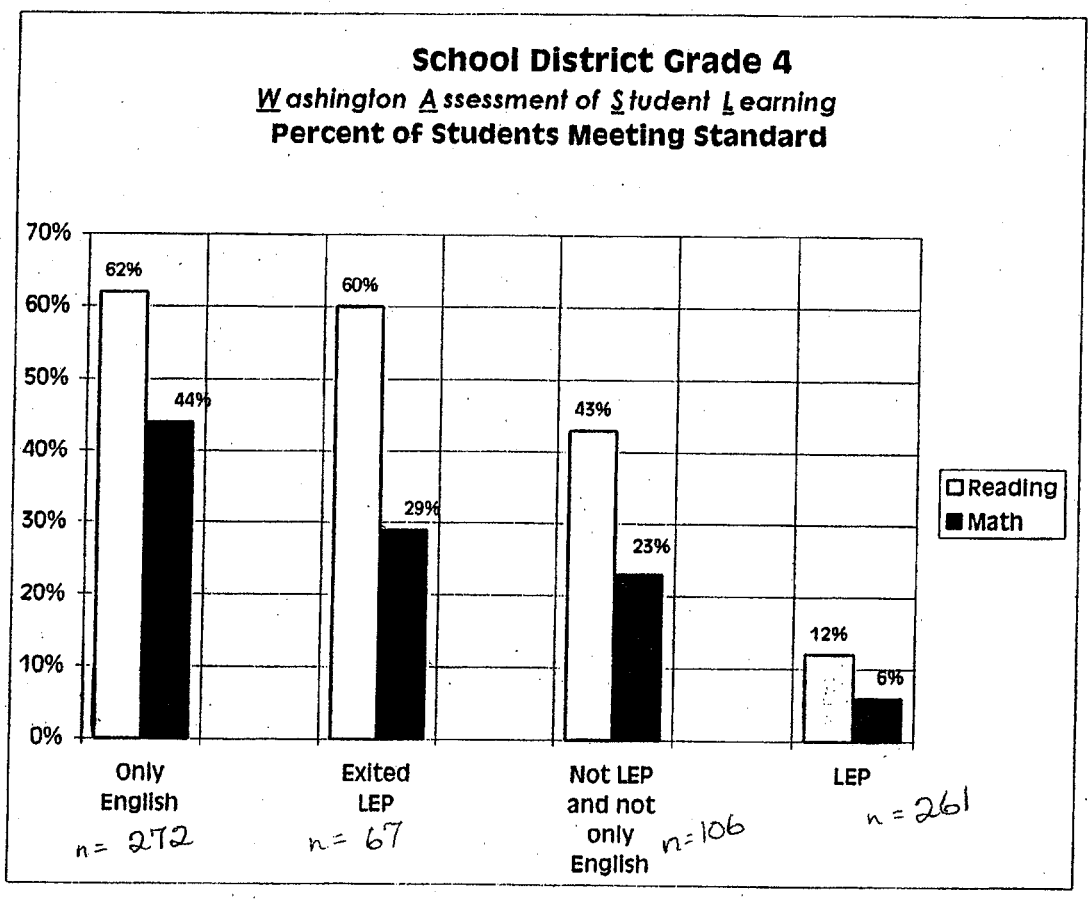
Appendix G

Longitudinal WASL data, District B



Appendix H

2000 Grade 4 WASL data, District E



VITA

CLAUDIA ANN OLMSTEAD

GOALS

To work in teacher education and curriculum design
To further quality education for all students in K-12

EDUCATION

Ph.D., University of Washington, 2004, Education, Curriculum and Instruction. Specialization: Bilingual/ English as a Second Language. Cognates: Leadership and Policy Studies, Linguistics, Special Education

M. A., Eastern Washington University, 1993, Reading Education.

B. S., Northern Arizona University, 1976, Education. Major: Art, minor: Spanish.

EXPERIENCE

Instructor, Heritage University, 2004-, courses for bilingual/ ESL education

TOSA, Educational Service District 123, 2003-, Literacy/ ELL Education

ESL Specialist, Richland School District, 1990-2003.

Staff Consultant, University of Washington, Center for Instructional Development and Research (CIDR), 1998-1999

Outreach Assistant, University of Washington, Russian, East European and Central Asian Studies (REECAS), 1997-1998

Instructor, Columbia Basin College, 1986-1990, ESL/ Adult Literacy

ESL Teacher, Pasco School District, 1980-1986.

Arts and Crafts Teacher, St. Helens, OR, 1978-1980.

PUBLICATIONS

Journal writing for the second language learner: Two selected studies. (September, 1993) *Literacy Matters,* EWU, 11-13.