

## **High-Stakes Testing, Latinos, and English Language Learners: Lessons from Colorado**

Kathy Escamilla, Elizabeth Mahon,  
Heather Riley-Bernal, and David Rutledge  
University of Colorado, Boulder

### **Abstract**

In Colorado, the Colorado Student Assessment Program (CSAP) has been created as the performance standard to determine progress that Colorado students are making toward meeting content standards. This study utilized results of the CSAP across 3 years (1999–2001) to determine the impact that standards-based education in Colorado is having on Latino students in general, and on Latino English language learners (ELLs) specifically. CSAP results in reading and writing in English and Spanish were compared for Latinos taking the CSAP in Spanish, Latinos taking the CSAP in English, and all Colorado third and fourth graders. This study also examined the extent to which school report card grades were affected in schools with large numbers of ELLs. Results indicate that the percentage of Latinos meeting state standards as measured by the Spanish CSAP is equivalent to, and in some cases higher, than the percentage of Latinos who are taking the CSAP in English. However, a gap exists between Latinos, no matter what their language of instruction and testing, and all Colorado third and fourth graders. This study found that school report card grades are lower in schools with large numbers of ELLs; this raises questions about the negative impact of school report card grades on schools that have high numbers of ELLs.

The sprinter and high jumper each concentrate on one event and may excel in it. The hurdler concentrates on two different skills, sprinting and high jumping, trying to combine a high standard in both. With only a few exceptions, the hurdler will be unable to sprint as fast as the sprinter or jump as high as the high jumper. This is not to say that the hurdler is a worse athlete than the other two. Any comparison of the two makes little sense. This analogy suggests that comparing the language proficiency of a monolingual with a bilingual's dual language or multilingual proficiency is similarly unjust (Baker, 2001, p. 8).

We need to look at bilingual schools as educating students to be hurdlers, capable of qualitatively different skills than sprinters or high jumpers. As of yet, state assessment systems fail to acknowledge this diverse type of learning.

## Introduction

In 1994, the Goals 2000 Educate America Act was enacted, creating standards-based educational reform that set challenging standards for all students (McLaughlin & Shepard, 1995). Since then, 49 of 50 states have established standards-based education reforms (Cunningham, 2000). In the majority of states where standards-based education programs have been implemented, they have focused on two components—content and performance. Content standards establish the substance of what students should learn during the course of their K–12 schooling, and performance standards establish the ways in which the attainment of standards will be measured (McLaughlin & Shepard, 1995). From its inception, the rhetoric of standards-based educational reform has emphasized the inclusion of *all* students into these reform efforts. *All* was meant to include students who are culturally and linguistically diverse as well as students with exceptional needs.

Recently, numerous questions and concerns have been raised about the application of standards-based education reforms to students who are English language learners (ELLs) (August & Hakuta, 1997; Gottlieb, 2001; McLaughlin & Shepard, 1995; Menke, 2001). Little or no controversy has surfaced with regard to content standards for ELLs. Indeed, there is widespread agreement that ELLs can and should meet challenging content standards. However, there is a great deal of concern about how the performance standards have been applied to all students, most especially to ELLs.

Performance standards define how students will demonstrate knowledge and/or skills, and the mastery of the content standards. For most states, performance standards have become equated with a single state-mandated test administered on an annual basis. On these annual performance tests, all students are expected to demonstrate their knowledge and skills of the content standards. In addition, in many states, results of these annual tests are the primary criteria used to assign report card grades to individual schools. This is the case in Colorado, where the Colorado Student Assessment Program (CSAP) test is used to measure student mastery of state content standards and is also used as the primary criterion for assigning school report card grades.

This study was undertaken to begin to assess the impact that the CSAP tests in reading and writing are having on Latino students in general, and more specifically on Spanish-speaking Latinos in elementary schools in Colorado. This study presents findings from 3 years of data on the performance of third and fourth graders on the Spanish and English CSAP, and examines

the influence of high numbers of ELLs on school report card grades in eight school districts. The study considers Colorado as a case study that contributes to the growing concerns about how standards-based education reforms, particularly performance standards, have been applied to culturally and linguistically diverse students. These concerns are exacerbated when performance standards are reduced to the use of a single assessment measure in an atmosphere of high-stakes testing.

### **Need for the Study and Research Questions**

Standards-based education reforms in Colorado were legislated in several phases. The passage of House Bill 93–1313 (Colorado State Legislature, 1994a) and House Bill 97–1249 (Colorado State Legislature, 1994b) created the mandate for school districts to develop content standards and for the state to create performance-based assessments to measure the attainment of content standards. Performance-based assessment involved the creation of the CSAP, which was implemented for the first time in 1999. In 1999, the state passed Senate Bill 186, which created the mandate for school accreditation and report cards utilizing the annual CSAP results as the primary measure for assigning school report card grades.

Senate Bill 186 provided two accommodations for students in Colorado who are labeled ELLs. These included exemptions from English-language CSAP testing for 3 years and the development of CSAP assessments in Spanish for Grades 3 and 4 in the areas of reading and writing (Colorado State Legislature, 2000). Spanish and English CSAP assessments in the areas of reading comprehension for third graders and reading and writing for fourth graders were administered for the first time in the spring 1999 and have been given annually since. These accommodations mirror those that other states provide for linguistically diverse students (Rivera, Stansfield, Scialdone, & Sharkey, 2000).

The need for this study arose early during the 1999–2000 school year when researchers and practitioners in Colorado noted that there was little accountability with regard to the progress that ELLs were making toward meeting state standards. This lack of accountability became evident when the Colorado Department of Education (CDE) delivered its annual report to the state legislature in January 2000. This report did not have a single reference to the plight or progress of ELLs in the state (Colorado Department of Education, 2000). First-year CSAP results were aggregated and disaggregated in multiple ways. They were disaggregated by gender, ethnicity, 13 different types of handicapping conditions, and socioeconomic status. Notably absent was any information on the results for students taking the CSAP assessment in Spanish, and no data were presented that had been disaggregated with regard to language proficiency and English CSAP results.

This study was undertaken for the purpose of examining the impact that the CSAP assessment system is having on ELLs and Latino students in the

state of Colorado. A second purpose was to examine the impact that large numbers of ELLs in Colorado schools might have on school report card grades. It is important for state accountability systems to recognize the impact of standards-based reforms on linguistically and culturally diverse students and the schools that serve them. Data analyzed and reported in this study are part of a larger study on the impact of the CSAP on ELLs (Escamilla, Aragon, Grassi, Riley-Bernal, Rutledge, & Walker, 2000; Escamilla, Mahon, Riley-Bernal, & Rutledge, 2001).

The analysis of third- and fourth-grade CSAP results provides the focus of this study for several reasons. First, CSAP results for third and fourth graders are available for 3 years, thereby making it possible to look at impact on students across time. Other CSAP assessments have only begun to be administered over the past 1 or 2 years. Second, CSAP assessments in Spanish reading and writing are only available at these grade levels. Therefore, it is only at these levels that testing accommodations such as native language assessment and its potential impact can be assessed. Finally, 95% of the bilingual education or dual language programs in Colorado are situated in elementary schools, and 98% of school districts in Colorado serving ELLs identify Spanish as the language spoken by the majority of ELLs (Escamilla et al., 2000, 2001). Research questions for this study are as follows:

1. At the third-grade level, how do CSAP results compare for Latino students taking the Spanish CSAP, Latinos taking the English CSAP, and all Colorado third-grade students taking the English CSAP?
2. At the fourth-grade level, how do CSAP results compare for Latino students taking the Spanish CSAP, Latinos taking the English CSAP, and all Colorado fourth graders taking the English CSAP?
3. How have school report card ratings been influenced in schools with large percentages of ELLs?

### **Profile of the State of Colorado**

According to the CDE Web site, there were a total of 724,508 public school (K–12) students in the state of Colorado in fall 2000. Of these students, 159,600 (22%) were classified ethnically as Latino with about 65% of the Latino population being of Mexican or Mexican American descent. The state has annual reporting requirements that include gathering data on ethnic groups, gender, socioeconomic status, and numbers of students in special education programs. However, efforts to gather data about the number of ELLs in the state have been less than comprehensive. Since 1980, the state has gathered information only on ELLs who qualify for state funds through the English Language Proficiency Act (Colorado Department of Education, English Language Acquisition Unit, 2000), or for school districts receiving Title VII funds from the federal government. The English Language Proficiency Act provides funds for schools to serve ELLs for 2 years. However, applying for

these funds is voluntary on the part of school districts, and only students eligible under the act are counted. Similarly, Title VII is a competitive federal grant program, and competing for funds is voluntary on the part of local school districts. As a result, the total number of ELLs in the state has never been fully documented. Data with regard to types of instructional programs available to serve ELLs in the state have been equally sketchy.

Without an accurate and detailed account of who the ELL population is and how it is being served, it is difficult to study the impact of the implementation of standards-based education reforms and the concomitant impact of the high-stakes performance system of the CSAP. For this reason, in 1999, six researchers at the University of Colorado at Boulder developed a survey to identify the numbers of ELLs in the state, how they were being identified, and how they were being served in Colorado schools (Escamilla et al., 2000). This survey was conducted via telephone communication with all 176 school districts in Colorado. The survey was conducted again in fall 2000 and fall 2001. Results from this survey provide important background data for this study and are briefly summarized below:

1. In fall 2000, there were a total of 57,692 students in Colorado identified as ELLs.
2. The number of ELLs in Colorado grew by 5,013 in 1 year (from 52,659 in 1999 to 57,672 in 2000). Data for fall 2001 are still being analyzed.
3. ELLs represent about 8% of all Colorado school students but are heavily concentrated in 12 metropolitan school districts and 6 mountain and rural school districts.
4. Eighteen school districts have 90% of the entire ELL population of the state.
5. Ninety-eight percent of Colorado school districts serving ELLs identify Spanish as the primary language spoken by ELLs.
6. About one third of all Latino students in Colorado speak Spanish as a first language.
7. In one large urban school district, ELLs constitute 28% of the total student body (over 17,000 students).

Data summarized above demonstrate that the state has a large and rapidly growing number of ELLs, and that these students are heavily concentrated in a few school districts. This uneven distribution of ELLs in Colorado's school districts and schools may mean that high-stakes testing impacts some districts in different ways than it affects others. Further, the vast majority of ELLs in Colorado speak a common native language, Spanish, and are labeled ethnically as Latino. In order to begin to understand how standards-based education reforms and CSAP assessment testing are impacting these students and their schools, it is imperative to disaggregate and examine test results for this population.

## **Colorado Student Assessment Program (CSAP)**

The CSAP was developed by CTB/McGraw-Hill (1998) as a means of assessing achievement of the Colorado content standards. The CSAP was intended to be a criterion-referenced test to measure students' mastery of content, not to compare students. The Spanish version of the CSAP was developed as a parallel assessment to the English CSAP. Both assessments were developed to measure the same content standards in reading and writing. Both have the same formats in reading and writing and a comparable number of items. Validity and reliability of both the English and Spanish CSAP tests were established using identical statistical measurements. Both the Spanish and English CSAP tests were field tested in Colorado schools. The English CSAP was field tested on third- and fourth-grade English-speaking students from a variety of ethnic and socioeconomic groups across the state. The Spanish CSAP was field tested on third- and fourth-grade students who were in bilingual education programs and learning to read and write in Spanish. For the purposes of this study, the Spanish and English CSAP tests for reading and writing in Grades 3 and 4 are considered to be comparable assessments.

Beginning in 1998–99, the CSAP was available in both English and Spanish in the areas of reading comprehension for Grade 3 and reading and writing for Grade 4. Results on the CSAP are reported as one of four categories of performance: unsatisfactory, partially proficient, proficient, and advanced. Students heretofore have taken the CSAP in either Spanish or English, but not both.

When the Colorado legislature mandated the development of English and Spanish CSAP tests, it intended for the English and Spanish CSAP tests to be used as parallel measures when determining the extent to which schools and school districts were meeting state reading and writing standards. However, the Spanish CSAP has not been given status or value equal to that of the English CSAP. Differential treatment between the Spanish and English CSAP tests and results have manifested in several ways. These include the treatment of data gathered on both tests, inequitable ways in which children and teachers are given information to help prepare for the test, and continued controversy over how Spanish CSAP results will be used by policy makers. Each of these issues is discussed in greater detail below.

Since the inception of CSAP testing in Colorado, Spanish CSAP results have been reported differently than English CSAP results. For example, all of the English CSAP results for this study were retrieved from the CDE Web site (<http://www.cde.state.co.us>). The Web site makes these data available to the public. Though the Spanish data are also public domain information, accessing the Spanish data has proven to be more of a challenge. To access the spring 1999 data, one of the researchers had to wade through boxes of information at the CDE. The data sat in the same boxes in which they had been turned in by school districts. In 1999, the data were neither summarized by the department, nor reported in any official documents or other outlets. For spring 2000, the

Spanish data were summarized by the CDE and placed in notebooks but were still not placed on the CDE Web site or available through other outlets. After much pressure, the spring 2001 data, for the first time, were summarized, placed alongside the English data on the CDE Web site, and reported to the state legislature.

In addition, teachers instructing children in reading and writing in Spanish have had fewer resources than those instructing in English to prepare students for the CSAP. In order to assist teachers in preparing students for the English CSAP, 25% of the reading and writing test items from the previous year's assessments are made available to them. Teachers may use these items to familiarize students with the CSAP test and to help them prepare. During the 3 years of this study, test preparation items were only available for the English CSAP. This has allowed English-speaking students the opportunity to practice authentic CSAP questions. No such items have ever been released for the Spanish CSAP, thereby placing teachers and students at a disadvantage because students taking the Spanish CSAP have not had the same opportunities to prepare for the test as English-speaking students have had.

During the 3 years of this study, there has been great controversy over the value of the Spanish CSAP tests. During the 1999 and 2000 school years, school districts were allowed to give students the Spanish CSAP tests; however, the CSAP tests were not counted when looking at a school and school district's overall progress toward meeting standards. Senate Bill 98, which passed in April 2001, mandated that Spanish CSAP tests be counted in a manner equivalent to English CSAP tests on a school's report card. The passage of Senate Bill 98 officially stated that Spanish CSAP results were to be given equal weight and status as English CSAP results. However, because the legislation was not enacted until 1 month after students had taken the CSAP tests for the 2000–01 school year, many districts were unsure as to whether to administer the CSAP in English or Spanish during that school year. Some districts opted for early exit to get ELLs ready to take the CSAP in English, while others opted to exempt ELLs from CSAP testing altogether until the controversy was resolved. Still others continued to prepare children to take the CSAP in Spanish. As a result, choices about whether to give ELLs the Spanish or English assessment CSAP were not made in a systematic or uniform way across the state; hence, these problems may have affected the reported results.

The results of this study, particularly with regard to the Spanish CSAP, need to be interpreted with the above caveats in mind. The differential treatment of the Spanish and English data points to the underlying tendency of the CDE to diminish and disregard the Spanish test results. Results on the Spanish CSAP are especially impressive given the number of obstacles that schools and school districts had to endure in order to administer the Spanish CSAP and to ensure that results on the Spanish CSAP would count toward a school's report card grade.

Finally, it is important to note that Colorado Senate Bill 186 (2000) provides for a 3-year exemption from English CSAP testing for students who are identified as ELLs, no matter what kind of instructional program they are in. While ELLs are expected to take the English CSAP, their scores are not counted on their school's accountability report. The English CSAP results, reported below, do not include scores of students who have been in Colorado schools for less than 3 years.

## Methods and Results

As stated previously, the study addressed three research questions. The questions and methods used to address them are discussed in detail below.

**Question 1:** At the third-grade level, how do CSAP results compare for Latino students taking the Spanish CSAP, Latinos taking the English CSAP, and all Colorado third graders taking the English CSAP?

The data used to address the first question were retrieved from the CDE Web site for English CSAP results (<http://www.cde.state.co.us>), and from files on the Spanish CSAP at the CDE. The data were examined at three levels: statewide, districtwide, and schoolwide for selected schools. The inclusion of the Spanish data contributes to a better understanding of the performance of Latinos as a group. Table 1 presents statewide results for third-grade students who took the Spanish and English CSAP in the spring of 1999, 2000, and 2001. Results are compared among Latinos taking the CSAP in Spanish, Latinos taking the CSAP in English, and all Colorado third graders taking the CSAP. Data are reported as the percentage of students from each group who scored at proficient or advanced levels on these tests.

Data presented in Table 1 contain three interesting findings. First, a greater percentage of Latinos taking the CSAP in Spanish scored at the proficient or advanced level than Latinos taking the CSAP in English. However, it is important to keep in mind that the sample sizes differ for each group, and so care should be taken when comparing the groups directly.

Second, results on the CSAP indicate a gap between Latinos (both those taking the test in English and those taking the test in Spanish) and all Colorado third graders in the percentage of students scoring at proficient or higher. It should be noted that over the past 3 years, the gap between Latinos taking the test in Spanish and all third graders has narrowed slightly, while the gap between Latinos taking the CSAP in English and all third graders has remained about the same.

Third, the percentage of students scoring at or above proficient has grown for all groups over 3 years. The addition of the Spanish CSAP data provides a more complete picture of Latino student achievement in Colorado.

However, an analysis of the overall state data does not tell the whole story of the achievement of ELLs. It is important to consider individual districts and schools. Over the past 3 years, some districts in the state have posted

Table 1

*A Comparison of Spanish and English CSAP Results for Third Graders—Reading, Spring 1999–2001*

	<b>Language of CSAP test</b>	<b>Percent at or above proficient</b>	<i>N</i>
<b>Spring 1999</b>			
Latinos taking Spanish CSAP <sup>a</sup>	Spanish	47	1,429
Latinos taking English CSAP <sup>a</sup>	English	46	Not reported by CDE
All Colorado third graders	English	67	52,780
<b>Spring 2000</b>			
Latinos taking Spanish CSAP <sup>a</sup>	Spanish	52	1,428
Latinos taking English CSAP <sup>a</sup>	English	49	10,741
All Colorado third graders	English	69	54,197
<b>Spring 2001</b>			
Latinos taking Spanish CSAP <sup>a</sup>	Spanish	56	1,795
Latinos taking English CSAP <sup>a</sup>	English	51	11,950
All Colorado third graders	English	72	55,207

<sup>a</sup>The authors are well aware that there are various labels used for defining this population (Hispanic, Latino, Chicano, etc.). The term Latino was chosen for this paper as an umbrella term to describe persons whose ancestors come from Latin America or the Caribbean and whose heritage language is Spanish. The state of Colorado uses the term Hispanic.

Spanish CSAP results that have exceeded the overall state CSAP results in English. Table 2 summarizes these results. In spring 1999, four school districts reported Spanish CSAP results that met or exceeded the statewide English results. In spring 2000, there were also four school districts whose Spanish CSAP results exceeded the statewide English results. In spring 2001, there were two school districts whose Spanish scores exceeded the English results. These individual district results provide data that suggest that some students taking the Spanish CSAP are meeting state standards in reading and writing at the third- and fourth-grade levels, and their scores are comparable to, or even higher than, the scores of some students taking the English CSAP.

Again, these data reflect the need to consider results on CSAP tests in Spanish along with results in English in order to gain a complete picture of the impact that standards-based education is having on Latino students in

Table 2

*Districts Where Spanish Third-Grade CSAP Results Met or Exceeded Statewide Third-Grade English Results, Spring 1999–2001*

<b>District</b>	<b>Spanish CSAP– percent at or above proficient</b>	<b>Statewide English CSAP–percent at or above proficient</b>
<b>Spring 1999</b>		
Adams County 14	76	67
Boulder	67	67
Eagle County	68	67
Ft. Lupton	68	67
<b>Spring 2000</b>		
Adams County 14	74	69
Brighton	71	69
Jefferson County	80	69
Roaring Fork	81	69
<b>Spring 2001</b>		
Brighton	92	72
Eagle County	76	72

Table 3

*A Statewide Comparison of Third-Grade CSAP Results in English and Spanish for Selected Denver Public Schools, Spring 1999–2001*

School	Spanish CSAP– percent at or above proficient			English CSAP– percent at or above proficient			District overall English CSAP– percent at or above proficient		
	1999	2000	2001	1999	2000	2001	1999	2000	2001
Bailey	36	43	X <sup>a</sup>	25	27	34	43	47	49
Bear Crest	50	X	X	52	67	40	43	47	49
Bo Williams	38	79	94	19	47	69	43	47	49
Castillo	52	39	40	31	36	34	43	47	49
Chavez	66	56	58	29	31	29	43	47	49
Coal	51	65	56	26	35	20	43	47	49
De la Puerta	53	X	X	38	29	27	43	47	49
Everest	45	58	56	49	30	48	43	47	49
Freemen	59	70	57	37	32	34	43	47	49
Gibbon	34	X	X	24	27	23	43	47	49
Gossamer	36	32	X	31	31	41	43	47	49
Knight	32	61	53	25	34	43	43	47	49
Moonbay	27	64	66	20	22	59	43	47	49
Reedsville	47	X	X	21	13	33	43	47	49
Smithers	57	66	67	34	42	46	43	47	49
Swallow	32	41	42	20	19	28	43	47	49
Vigil	62	50	81	14	10	28	43	47	49

*Note.* Pseudonyms are used for school names. Each of these schools is over 75% Latino. Therefore, Latinos comprise the majority of students taking English CSAP tests.

<sup>a</sup> Schools do not report data when there are 16 or fewer students tested.

individual school districts. This is particularly true for districts that have student populations that are majority Spanish speaking, and that have chosen to provide bilingual education as an educational option. The Spanish CSAP data indicate that these districts are accountable for Spanish-speaking students and are helping them meet state content standards in reading.

A similar pattern is detected when looking at individual schools. Adding the results for the Spanish CSAP presents a totally different picture for many individual schools. Failure to consider data on the Spanish CSAP when determining a school's report card grade presents a skewed picture of the achievement outcomes for many schools in Colorado with large numbers of Latinos and Spanish-speaking students. To illustrate the difference that adding Spanish CSAP results makes, Table 3 presents and compares results of Spanish and English CSAP tests for third-grade students in 17 elementary schools in a large Colorado school district. Each school is heavily impacted by ELLs, has an ethnic composition that is greater than 75% Latino, and is considered high priority for the English Language Acquisition department of the school district.

In spring 1999, there were 14 schools in this cluster that reported higher outcomes on the Spanish CSAP than on the English CSAP. That is, greater percentages of students taking the Spanish CSAP scored at or above proficient than students taking the English CSAP at the school and district level. In 2000, there were 13 schools that had higher CSAP scores in Spanish than English, and in 2001, there were 11 such schools. In addition, during these same years, there were many schools whose Spanish CSAP results exceeded English results districtwide. In 1999, there were 10 schools whose Spanish CSAP results exceeded the districtwide English scores, while in 2000 and 2001, there were 9 such schools. If data reports on these schools are limited to English CSAP results, overall results may be unfairly skewed in a negative direction and show that schools may be doing worse than they really are. The case of Vigil Elementary illustrates this point.

In October 1999, the *Denver Post* called for a restructuring of Vigil Elementary because of the school's abysmal results on the CSAP. In 1999, only 14% of the students at Vigil Elementary scored at or above proficient on the English CSAP. However, the *Denver Post* failed to look at the school's Spanish CSAP results, in which 62% of the students scored at or above proficient. The Spanish results exceeded the English results at this school and exceeded the district's overall percentage of students at or above proficient, which was 43%. Vigil Elementary has a student population with over 80% ELLs, and 98% Latinos, many of whom receive literacy instruction in Spanish. It seems only fair that an academic accountability for this school needs to include the Spanish CSAP results, as well as the English results. With 62% of the population scoring at or above proficient, Vigil becomes a star school in this district, rather than a school waiting for castigation from the state and the public.

In sum, through a lens of equity, there are two points that are clear when considering the third-grade CSAP data. First, there is a gap in achievement between Latinos and their Anglo counterparts even for English-speaking Latinos. Mere knowledge of English does not seem to narrow this gap. Second, students who are learning to read and write in Spanish are making progress in meeting the state's content standards, and their performance is improving across years and is on par with that of English-speaking Latinos.

**Question 2:** At the fourth-grade level, how do CSAP results compare for Latino students taking the Spanish CSAP, Latinos taking the English CSAP, and all Colorado fourth graders taking the English CSAP?

Data for the second question were retrieved from the CDE Web site for English CSAP results (<http://www.cde.state.co.us>), and from files at the CDE concerning Spanish CSAP results. Table 4 presents comparison data for fourth-grade reading and writing. Results are compared among Latinos taking the CSAP in Spanish, Latinos taking the CSAP in English, and all Colorado fourth graders taking the CSAP.

Results differed somewhat between the third and fourth grades. Furthermore, the addition of a writing assessment at fourth grade provides additional opportunities for data analysis. Fourth-grade results are summarized below.

Across the state, there is a decline in the percentage of students scoring at proficient or advanced between the third and fourth grades. This is true for all three comparison groups and has remained consistent across 3 years. The gap between both Latino groups (those taking the CSAP in English and those taking the CSAP in Spanish) was also apparent at fourth grade. However, it is important to note that in all three comparison groups, greater percentages of students are meeting state reading and writing standards each year.

In reading, a greater percentage of fourth-grade Latinos taking the English CSAP scored at or above proficient than Latinos taking the Spanish CSAP. Differences between the groups remained constant over the 3-year period, with Latino students taking the English CSAP continuing to perform better than Latinos taking the Spanish CSAP. This is in contrast to the third-grade results, in which greater percentages of Latinos taking the Spanish CSAP scored at or above proficient than Latinos taking the English CSAP.

In the area of writing, there was a large gap between all Colorado fourth graders and Latinos taking the English CSAP. There was also a gap between Latinos taking the English CSAP and Latinos taking the CSAP in Spanish. Latinos taking the CSAP in Spanish outperformed Latinos taking the CSAP in English. There was only a small gap between Latinos taking the CSAP in Spanish and all Colorado fourth graders.

For the writing assessment, the results of students taking the test in Spanish showed an interesting pattern. For all fourth graders and for the Latino students taking the CSAP in English, there is a significant gap between reading and writing, with both groups scoring higher in reading than in writing.

Table 4

*A Statewide Comparison of Spanish/English CSAP Results for Fourth Graders, Spring 1999–2001*

	<b>Language of CSAP test</b>	<b>Reading–percent at or above proficient</b>	<b>Writing–percent at or above proficient</b>	<b>N</b>
<b>Spring 1999</b>				
Latinos taking CSAP in Spanish	Spanish	23	27	1,076
Latinos taking CSAP in English	English	35	16	Not reported by CDE
All Colorado fourth graders	English	59	34	52,780
<b>Spring 2000</b>				
Latinos taking CSAP in Spanish	Spanish	28	30	1,104–Reading 1,099–Writing
Latinos taking CSAP in English	English	37	18	10,773 Reading & Writing
All Colorado fourth graders	English	60	36	54,827 Reading & Writing
<b>Spring 2001</b>				
Latinos taking CSAP in Spanish	Spanish	31	36	1,373–Reading 1,375–Writing
Latinos taking CSAP in English	English	40	18	11,837–Reading 11,833–Writing
All Colorado fourth graders	English	63	38	55,216–Reading 55,212–Writing

However, for Latinos taking the Spanish CSAP, this gap did not exist; writing outcomes were very similar to reading outcomes. Furthermore, for all 3 years, Spanish writing outcomes exceeded Spanish reading outcomes for these students. This result provides additional evidence that students learning in Spanish are on par with Latinos learning in English in meeting state reading and writing standards, and are ahead of English-speaking Latinos in meeting state writing standards. Again, simply knowing and learning in English does not seem to be boosting Latino student performance on state performance assessments. This pattern supports research that shows that the connection between reading and writing is different in Spanish than it is in English (Escamilla, 2001; Vernon & Ferreiro, 1999). More research is needed in this area to further analyze relationships between reading and writing across languages.

Similar to the third-grade analysis, the reading and writing results on the CSAP indicate a gap in achievement between fourth-grade Latinos (both those taking the test in English and those taking the test in Spanish) and all Colorado fourth graders. However, the gap is less significant between Latinos taking the Spanish writing CSAP and all fourth graders. In fact, results from the writing assessment in spring 2001 show that 36% of Latinos taking the writing assessment in Spanish scored at proficient or above, compared to 38% of all Colorado fourth graders. This represents the best indication of the closing performance gap between a minority group and all students. Unfortunately, this finding went completely unnoticed by the CDE and the Colorado media. A comparison between third- and fourth-grade reading results by year (Table 5) provides another opportunity to analyze these data.

Across the board, there is a drop in scores from the third grade to the fourth grade, both in a yearly analysis and in analysis by group. For example, looking at yearly data, a greater percentage of third graders scored at proficient or above than did fourth graders. This pattern remains consistent with all groups across 3 years. For example, in 1999, 47% of Latinos who took the third-grade reading CSAP in Spanish scored at or above proficient, whereas in the year 2000, this same general cohort took the fourth-grade CSAP, and only 28% scored at or above proficient. Similarly, results from Latinos taking the CSAP in English and all Colorado fourth graders showed a decline. The prevailing pattern is that all students statewide experienced a drop in scores, so this decrease in scores does not particularly stigmatize ELLs or their educational programs. The decreasing scores is a trend for all Colorado third and fourth graders.

It is also important to note that the gap between Latino students taking the CSAP in English and all Colorado fourth graders is greater in fourth grade than in third grade. Again, merely knowing English does not seem to improve performance on standards-based assessments among Latino students.

There are several potential factors that could explain the decrease in the number of students scoring at or above proficiency between the third and

Table 5

*A Comparison of Spanish CSAP Results for Third- and Fourth-Grade Reading, Spring 1999–2001*

	Third-grade reading		Fourth-grade reading	
	Percent	N	Percent	N
<b>Spring 1999</b>				
Latinos taking Spanish CSAP—at or above proficient	47	1,429	23	1,076
Latinos taking English CSAP—at or above proficient	46	Not reported	35	Not reported
Statewide results—at or above proficient on English CSAP	67	52,780	59	52,780
<b>Spring 2000</b>				
Latinos taking Spanish CSAP—at or above proficient	52	1,428	28	1,104
Latinos taking English CSAP—at or above proficient	49	10,741	37	10,773
Statewide results—at or above proficient on English CSAP	69	54,197	60	54,827
<b>Spring 2001</b>				
Latinos taking Spanish CSAP—at or above proficient	56	1,795	31	1,373
Latinos taking English CSAP—at or above proficient	51	11,950	40	11,837
Statewide results—at or above proficient on English CSAP	72	55,207	63	55,216

fourth grades. It has been well documented that in third grade, students are learning to read, whereas in fourth grade, students are reading to learn. It is also important to note that the Colorado reading and writing content standards for the fourth grade are considerably more difficult than for the third grade, and the CSAP test has twice as many items (CTB/McGraw-Hill,1998). Both the curriculum and the CSAP test become more difficult and cognitively demanding at the fourth-grade level. Furthermore, the third-grade CSAP in reading measures only two standards, whereas the fourth-grade CSAP measures all five reading standards. This increase in breadth of the test makes the fourth-grade test more difficult for students. There are multiple explanations for declining achievement levels between the third and fourth graders, but for the purposes of this study, it is important to note that lower results occur for all groups of students in the study and not just for ELLs. Patterns for students taking the CSAP in Spanish are the same as for Latinos taking the English CSAP and for all Colorado fourth graders.

However, it is also important to note that for ELLs, fourth grade is often the year that they are expected to transition from Spanish to English reading. Hence, fourth-grade data should be interpreted in light of the possibility that less emphasis is placed on learning to read and write in Spanish in the fourth grade, with more emphasis on learning to read and write in English. Also, the state 3-year exemption runs out for students in their fourth-grade year, so results for Latinos on the English CSAP reflect former ELLs who are now taking the test in English instead of in Spanish. More research is needed to document what the shift from Spanish to English CSAP testing is doing to students in the fourth grade.

**Question 3:** How have school report card ratings been influenced in schools with large percentages of ELLs?

To address this research question, data from the CDE regarding school report card grades were analyzed. These data were published on September 14, 2001, in the *Denver Post*. These data were compared to and combined with data on numbers of ELLs reported by school districts. Individual school report card grades are greatly impacted by having high numbers of ELLs in a given school. The impact is significant despite the 3-year exemption of ELLs from English CSAP tests. English CSAP data presented below do not include recent arrivals or students who are new to learning English. These students' results on English CSAP tests were excluded from school report card calculations. Data include only students who are identified as ELLs and who have been in Colorado schools more than 3 years. They also include results on Spanish CSAP tests for these schools.

In order to examine how numbers of ELLs may influence school report card grades, the researchers contacted eight school districts in the state that have large numbers of ELLs. Districts were asked to identify elementary schools that were "heavily impacted" by ELLs. Each district had a different definition for the term "heavily impacted." Some districts, like Denver Public Schools,

consider schools to be heavily impacted if they have 60 or more ELLs. Other districts, such as Aurora School District, consider schools to be heavily impacted if 33% of the students in the school are ELLs. Other districts use multiple criteria for considering a school to be highly impacted. For example, districts may consider training of staff, available models of instruction, services of support staff, history of the particular school, and achievement level of ELLs. Some districts include in the high-impact category any school that receives services from an ESL or bilingual teacher. The broader definition increases the number of schools considered high impact.

The criteria for high-impact schools were too varied from school to school to draw any reasonable conclusions. Therefore, after much discussion, we settled on a definition of “highly impacted” that included schools where 20 or more students were considered to be ELLs. This definition was the most inclusive, as it included all local district definitions of the term “highly impacted.” We divided schools into two categories: low impact with regard to ELLs and high impact with regard to ELLs. We then calculated the percentage of low- and high-impact schools that were in each report card category (unsatisfactory, low, average, high, excellent). Table 6 below presents the number and percentage of schools highly impacted by ELLs in these eight districts, and Table 7 presents a cross tabulation of high-impact and low-impact schools, with respect to ELLs, and their report card grades.

Table 6 indicates that in these eight districts, there are 367 elementary schools that received school report card grades, with 127 of these schools identified as “heavily impacted” by ELLs. It should be noted that schools highly impacted by ELLs account for 33% of all elementary schools in these eight districts. At first glance, it would appear that the law exempting ELLs from the CSAP test would help the overall report card grades of these schools. However, this does not seem to be the case. As Table 7 indicates, there is a significant relationship between a school’s report card grade and the number of ELLs in the school. For the districts sampled, 76.5% of the unsatisfactory schools were highly impacted by ELLs while only 4.5% of the heavily impacted schools received grades of excellent. Furthermore, 62% of the schools receiving low grades were highly impacted by ELLs. It must be noted that schools that report having significant numbers of ELLs also have significant numbers of former ELLs, and are also considered to be economically disadvantaged or low-income schools.

A clear way to see how high-impact schools are more likely to receive a low or unsatisfactory grade is presented in Figure 1. It demonstrates that schools that are highly impacted by ELLs are much more likely to have report card grades in the low or unsatisfactory range. It is important to note that students included in English CSAP testing in these schools are not newly arrived ELLs. Newly arrived ELLs are exempted from English CSAP testing for 3 years. The numbers reported above represent students tested on the English

Table 6

*Number and Percentage of Schools Highly Impacted by ELLs—Eight Selected School Districts, 2000–2001 School Year*

<b>District</b>	<b>Number of elementary schools in district<sup>a</sup></b>	<b>Number &amp; percent of schools with high ELL impact</b>
Danford	90	48 (53%)
Arbor	39	13 (33%)
Johnson County	104	26 (24%)
Berry Vine	36	9 (25%)
Pewter	29	4 (14%)
Arlene County #50	16	4 (25%)
St. Vincent	19	11 (58%)
Crescent Crest	34	12 (35%)
<b>Total</b>	<b>367</b>	<b>127</b>

*Note.* Includes elementary school data only. Only schools where 20 or more students were classified as ELL were included

<sup>a</sup>Includes only schools that received report card grades from the Colorado Department of Education

Table 7

*Cross Tabulation of High- and Low-Impact ELL Schools and School Report Card Grades, 2000–2001 School Year*

<b>ELL impact</b>	<b>Unsatisfactory</b>	<b>Low</b>	<b>Average</b>	<b>High</b>	<b>Excellent</b>
High impact	76.5%	62%	28.3%	13%	4.5%
Low impact	23.5%	38%	71.7%	87%	95.5%

*Note.* Includes elementary school data only. Includes only schools that received report card grades from the Colorado Department of Education

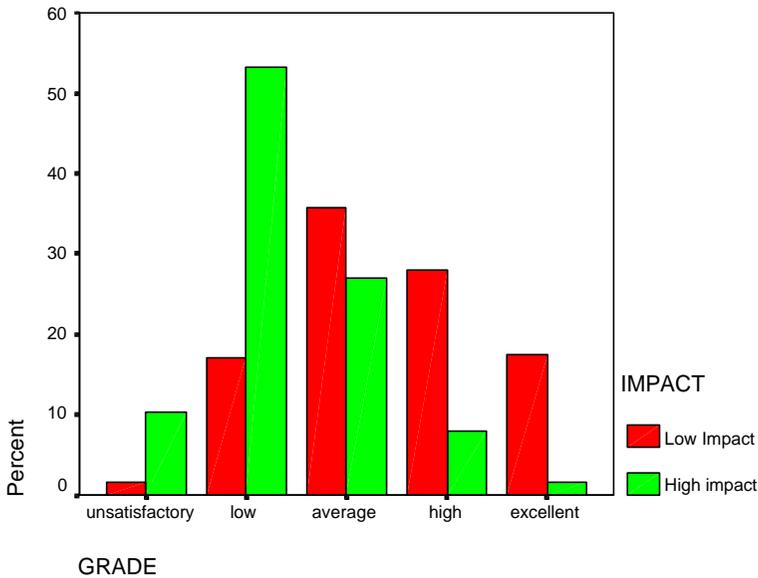


Figure 1. Frequency of grades for both high-impact and low-impact schools in the eight school districts  $N = 367$ .

CSAP who have run out of their 3-year exemptions. These data document that even with a 3-year exemption, schools are still negatively impacted in the calculation of school report card grades if they have large numbers of ELLs. A question arising from these data and other studies (Gottlieb, 2001; Rivera, Stansfield, Scialdone, & Sharkey, 2000) is at what point does the English CSAP become a valid measure to determine whether ELLs are meeting state standards?

In some aspects, this type of analysis is similar to looking at school grades for schools with high percentages of students receiving free and reduced lunch (McQuillan & Englert, in press). Wealthier schools receive higher grades, while economically poorer schools receive lower grades. In this case, schools with large numbers of ELLs received lower grades than schools with fewer ELLs. It should be noted that neither the socioeconomic status nor the impact of ELLs on schools was reported in this round of report card grades, though both variables clearly have an impact on ratings. It should also be noted that schools with large numbers of ELLs and Latino students are generally also economically disadvantaged schools.

There are many implications from the school report card data. The first set of implications addresses accountability and exemption issues. The school report card data raise questions about the quality of instructional programs for ELLs, and whether or not there is sufficient accountability for their progress before they take the English CSAP. The fact that ELL scores on the English CSAP do not count for 3 years does not seem to be positively impacting the

school ratings. The intent of the 3-year exclusion from the English CSAP was that neither children nor schools would be penalized. The good intentions of the policy makers do not appear to be having an impact on the outcomes. These data also imply that it takes longer than 3 years to reach full proficiency in English. While there is a plethora of research to support this observation (e.g., Cummins, 1986; Collier & Thomas, 1995; Mitchell, 1997), additional research specifically related to English CSAP outcomes and ELLs is needed.

A second set of implications relates to the use of CSAP data to rank schools. In Colorado, as in other states, the CSAP and school report card grades may create an unwelcoming environment for ELLs. If having large numbers of ELLs in a school stigmatizes a school as low or unsatisfactory, then there is little incentive to work or teach there. There is little reason for students themselves to take pride in such a school. Furthermore, attitudes toward ELLs and their families are likely to get worse, not better. Ruiz (1988) outlines three basic orientations in language planning: language as problem, language as resource, and language as right. The high-stakes testing environment places schools in a position to develop a language-as-problem orientation because students who are not native speakers of English may bring down school report card grades. The stakes are so high in low-performing schools, and the pressure to improve might force schools to look for quick fixes to improve test scores, rather than long-term solutions to improve educational programs for students.

During the 2000–01 school year, Spanish CSAP results were included in the criteria for determining a school’s report card rating. A study released by the CDE (2001) reported that including the Spanish CSAP results in school report card ratings was a zero-sum game. That is, inclusion of these scores caused ratings in some schools to improve (nine schools improved) and caused other ratings to decline (nine total declined). We submit that it is beneficial to include Spanish CSAP results in school report card ratings because this inclusion provides a better picture of an entire school’s progress toward meeting state standards, and it includes a greater number of students in the accountability process.

### **Summary and Implications**

This study was motivated by the need to address the impact of the standards-based education movement on ELLs in Colorado. Consistent with national trends, the state of Colorado utilizes two major accommodations in an effort to include ELLs in standards-based assessment mandates (Rivera, Stansfield, Scialdone, & Sharkey, 2000). These include providing assessment in students’ native languages and providing exemption from English-language testing for several years. In the case of Colorado, students may take CSAP assessments in Spanish in the areas of reading and writing for Grades 3 and 4. In addition, while all students must take the English CSAP assessments, the

results of ELLs are not counted for 3 years, thus providing the equivalent of a 3-year exemption from English CSAP tests for ELLs.

Results of the study documented a large gap in performance between Latinos taking the CSAP assessments in Spanish, Latinos taking the CSAP in English, and all Colorado third and fourth graders. By disaggregating data between English-speaking and Spanish-speaking Latinos, this study documented that the achievement gap cannot be attributed to language alone. The achievement gap exists for both English- and Spanish-speaking Latinos. Furthermore, this gap is wider for reading at the third-grade level and for writing at the fourth-grade level for Latinos taking the CSAP in English than for Latinos taking the CSAP in Spanish. In short, knowledge of English does not seem to be helping Latinos better meet state content standards or close the achievement gap.

The purpose of the CSAP assessment program is to determine the extent to which schoolchildren in Colorado are meeting state content standards. With regard to third-grade reading and fourth-grade writing, results of this study indicate that students taking the CSAP in Spanish are doing better than Latinos taking the CSAP in English. This finding is surprising given the lack of emphasis that the state has placed on the Spanish CSAP, and the contradictory and controversial policies issued from the CDE with regard to whether or not the CSAP was going to count on school report cards. This finding is even more surprising given the enormous pressure on school districts to teach in English and to transition students from Spanish-language instruction to English-language instruction as quickly as possible. In short, Spanish CSAP results indicate that Spanish-speaking Latinos are on par with English-speaking Latinos in meeting state reading and writing standards in the third and fourth grade in spite of great pressure to limit teaching in Spanish. Spanish speakers taking the Spanish CSAP do as well and in some cases better than English-speaking Latinos taking the English CSAP. There is a need to follow Spanish speakers who take the CSAP in Spanish and to monitor their progress as they begin to take the CSAP in English to determine if there is a positive correlation between Spanish and English outcomes.

In addition, utilization of Spanish CSAP results offers a more complete picture of school and district performance. If standards-based assessment is truly targeted at all children, then the test results for all children must be included in school report cards and ratings. As documented in this study, failure to include Spanish CSAP results negatively skews the profile of student academic achievement at schools that have large numbers of Spanish-speaking ELLs who are learning to read and write in Spanish.

Study results indicate that schools with large numbers of ELLs are negatively impacted on school report card ratings despite the 3-year exemption for ELLs. Using exemptions as an accommodation for ELLs does not seem to be helping school ratings. Furthermore, exempting students from the English CSAP for 3 years may not be helping ELLs meet state standards. There is

currently no accountability mechanism for assessing ELLs' progress toward meeting state content standards during the time that these students are exempted from the English CSAP. ELLs may be acquiring a great deal of English in Colorado schools, no matter what the instructional program; however, 3 years may not be a sufficient amount of time to acquire enough English to meet state content standards in reading and writing. In addition, allowable exemptions may mean that ELLs are not given the same quality of instructional programs as other students, even in highly impacted schools, because their scores on the CSAP will not count for 3 years.

This study has documented that a pattern exists between schools that are highly impacted by ELLs and low and unsatisfactory ratings on school report cards. Further research is needed into why this relationship exists and to explore potential solutions to this dilemma through policy changes, improved instructional programs, or both. For now, it seems as if the current CSAP testing and reporting system has been designed to punish, rather than support, schools and school districts with large numbers of ELLs. Clearly, some schools and districts face greater challenges than others. They need additional resources and more attention, not castigation.

The focus of this study was to examine how standards-based education reforms in Colorado are impacting Latino students, both those who speak Spanish and those who speak English. We used the CSAP to address the research questions. However, it is important to discuss the limitations of assessment data in addressing these questions. With regard to future research in this area, it is not sufficient to discuss the impact that standards-based education reforms have had on Latinos and ELLs without also discussing opportunities to learn (see McLaughlin & Shepard, 1995). It is clear that content standards are directed at all students in Colorado and that performance standards continue to indicate gaps in achievement between both English- and Spanish-speaking Latinos and other children. It is also clear that utilizing performance-based assessments in two languages, as in Colorado, adds additional information to the discussion on the impact that standards-based education is having on ELLs.

McLaughlin and Shepard (1995) argue that if standards-based education reforms are to truly impact all students, an additional standard must be added, that of opportunities to learn. Opportunity-to-learn standards focus on the resources, programs, and quality of teachers and educational facilities needed to enable all students to meet the content and performance standards. Opportunity-to-learn standards are critical for improving educational opportunities for Latino students in general, but most especially for ELLs. Sadly, opportunity-to-learn standards have received little attention in Colorado during the past 6 years of standards-based reform efforts.

We began this paper with the analogy of bilingual students as unique athletes—hurdlers—whose success depends on their ability to both sprint and jump. Similarly, bilingual students' success depends on their skills in both Spanish and English, and these skills should be assessed and recognized

accordingly rather than represented by measures focused only on one language or the other. Overall, what is important to both policy and practice is that educators need to recognize what students *can* do rather than focusing on what they cannot do. As long as state standardized tests only assess hurdlers' sprinting or high-jumping skills, they fail to evaluate and recognize all that multilingual students can do.

## References

- August, D., & Hakuta, K. (Eds.). (1997). *Improving schools for language minority children*. Washington, DC: National Academy Press.
- Baker, C. (2001). Bilingualism: Definitions and distinctions. In C. Baker, *Foundations of bilingual education and bilingualism* (pp. 8–9). Tonnawanda, NY: Multilingual Matters.
- Biliteracy in two languages [Editorial]. (1999, Oct. 14). *Denver Post*, p. 6B.
- Collier, V., & Thomas, W. (1995). *Language minority student achievement and program effectiveness*. Washington, DC: National Clearinghouse for Bilingual Education.
- Colorado Department of Education. (2000). *Colorado student assessment program: Student performance in reading comprehension grade 3, reading and writing grade 4, and reading and writing grade 7: Annual report to the Colorado General Assembly*. Denver, CO: Author.
- Colorado Department of Education. (2001). *Taskforce report on students whose dominant language is not English*. Denver, CO: Author.
- Colorado Department of Education. (2001). *Historic reading, writing, and math CSAP: 1997–2001 state and district summary*. Retrieved February 17, 2003, from [http://www.cde.state.co.us/cdeassesses/as\\_Historiccsap.htm](http://www.cde.state.co.us/cdeassesses/as_Historiccsap.htm)
- Colorado Department of Education, English Language Acquisition Unit. (2000). *ELL students by district 2000*. Retrieved February 17, 2003 <http://www.cde.state.co.us/cde-english/elau-elpa.htm>
- Colorado State Legislature. (1994a). H.B. 93–1313.
- Colorado State Legislature. (1994b). H.B. 97–1249.
- Colorado State Legislature. (2000). S.B. 186.
- Colorado State Legislature. (2001). S.B. 98.
- Colorado schools accountability report: 2000–01 school year. (2001, September 14). *Denver Post*, pp. F16–F17.
- CTB/McGraw-Hill. (1998). *Colorado Student Assessment Program (CSAP): Final technical report*. Monterey, CA: Author.
- Cummins, J. (1986). *Empowering language minority students*. Sacramento: California Association for Bilingual Education.

- Escamilla, K. (2001). Bilingual means two: Assessment issues, early literacy and Spanish speaking children. *Reading Research Symposium for Second Language Learners* (pp. 1–16). Washington, DC: National Clearinghouse for Bilingual Education.
- Escamilla, K., Aragon, L., Grassi, E., Riley-Bernal, H., Rutledge, D., & Walker, D. (2000, April). *Limited English proficient students and the Colorado Student Assessment Program (CSAP): The state of the state*. Denver: Colorado Association for Bilingual Education.
- Escamilla, K., Mahon, E., Riley-Bernal, H., & Rutledge, D. (2001). *Limited English proficient students and the Colorado Student Assessment Program (CSAP): The state of the state, Year II*. Denver: Colorado Association for Bilingual Education.
- Gottlieb, M. (2001). Four “A’s” needed for successful standards-based assessment and accountability. *NABE News*, 24(6), 8–12.
- McLaughlin, M., & Shepard, L. (1995). *Improving education through standards-based reform*. Washington, DC: National Academy of Education.
- McQuillan, P., & Englert, K. (in press). The return to neighborhood schools, concentrated poverty, and educational opportunity: An agenda for reform. *Hastings Constitutional Law Quarterly*.
- Menke, K. (2001). When all means all: Standards based education reform and English language learners. *NABE News*, 24(5), 4–7.
- Mitchell, D. (1997). New insights into bilingual education: Learning English takes longer than commonly thought. *Fiat Lux*, 7(4), 7–10.
- Rivera, C., Stansfield, C., Scialdone, L., & Sharkey, M. (2000). *An analysis of state policies for the inclusion and accommodation of English language learners in state assessment programs during 1998–1999*. Arlington, VA: George Washington University, Center for Equity and Excellence in Education.
- Ruiz, R. (1988). Orientations in language planning. In S. McKay & S. Wong (Eds.), *Language diversity: Problem or resource?* (pp. 3–25). Cambridge, MA: Newbury House.
- School report cards out, generate mixed reviews. (2001, September 14). *Denver Post*, pp. 1–2B.
- Vernon, S., & Ferreiro, E. (1999). Writing development: A neglected variable in the consideration of phonological awareness. *Harvard Educational Review*, 69(4), 395–415.

