

Defining and Documenting Success for Bilingual Learners: A Collective Case Study

María E. Torres-Guzmán
Columbia University

Jorgelina Abbate and María Estela Brisk
Boston College

Liliana Minaya-Rowe
University of Connecticut

Abstract

This article examines the difficulties inherent to measuring bilingual program success and the need for broader and fairer assessment strategies for bilingual students. Drawing from our collective case study, we confirm that there are significant data sources available and accessible to the schools/programs but that their formats are not easily comprehensible for schools attempting to showcase their programs. We also report how the collection and compilation of assessment is primarily in the hands of the school administrators and, thus, may not be efficiently used for the improvement of teaching and learning. Despite the difficulties of data gathering and the shortcomings on the use of information, we suggest that in the schools studied, the evidence we gathered supported their perspectives on success.

Introduction

The national debates on bilingual education have called upon the field to focus more closely on issues related to success and failure. Experts in the field of bilingual education have demonstrated the benefits of using native language instruction with students for whom English is not the first language (Greene, 1997; Ramirez, Yuen, & Ramey, 1991; Thomas & Collier, 1998; Willig, 1985, 1987). However, these benefits have not been translated convincingly to the public. We have witnessed two states, California and Arizona, overturn bilingual education through the election polls. While the

overwhelming evidence suggests that the problem is not at the educational but at the political level, the reaction of the public warrants a closer examination of the definitions of success and the evidence presented. This may provide insights as to how such differences in views about bilingual education are constructed (Tse, 2001).

Portraits of Success (PoS) is a project that seeks to develop and maintain a database of successful bilingual programs. Success is determined by a combination of components, but it continues to pay attention to evidence of student achievement (Brisk, 2000). The process of selection for showcasing a school is rigorous; the application is detailed and requires that schools present student outcome data to prove accomplishment of their goals. Throughout our work with PoS, we have found that schools face real difficulties when asked to provide evidence to showcase their programs. They submit thorough program descriptions, basic statistics related to the student population served, and evidence of positive community response. The area that remains problematic is reporting student outcomes. Given the increasing implementation of state-mandated curriculum frameworks and high stakes student assessment in the last few years, the unavailability of student outcome data may sound surprising.

The schools' ongoing challenge to report student outcomes suggests that the absence or poor implementation of state or district policies often cause the assessment of bilingual students to be overlooked, or to be administered with inappropriate assessment tools. In this collective case study within three schools in three different urban areas of the eastern coast we explored the process of evidence gathering through the following questions:

1. What outcome data are available for bilingual students in bilingual programs/schools considered successful? How accessible are the data?
2. What do the available data say about student performance?
3. How do staff associated with those programs construct their statements of success?

The case study schools were reputed to offer quality bilingual education but had not yet submitted an application to PoS. We were interested in understanding not only the difficulties in gathering data but also how the constructions of success would stand in relation to theory when we critically examined the evidence gathered. Additionally, while we were concerned about specific measures of success associated with PoS that guided us in obtaining the data, we also sought to gain a broader sense of how school personnel assessed the progress of their students, how they perceived the importance of student outcomes in guiding further instruction, and the evidence they used to support their view of the program's success.

In this article, we highlight the processes of identifying the data available and issues of accessibility that arose in the data gathering process. We center the discussion of our findings on the differences in interpretation of data arising from

competing student assessment perspectives and how the differences relate to issues of added value and educational equity for linguistically diverse populations.

Assessing the Effectiveness of Bilingual Education

Suarez-Orozco and Suarez-Orozco (2000) propose that urban school districts need to implement general assessment systems that take into account students' academic and linguistic development, given that one out of six children lives in a household headed by an immigrant and in which a language other than English is spoken. In other words, the relationship between linguistic and academic development is not solely the purview of bilingual education because not all communities choose to service their students in the same way. Moran & Hakuta (1995) point to the significance of community contexts with respect to program designs. They state that bilingual program models are responses;

to varied populations as well as to the political, social, and educational objectives of different school sites. Communities differ not only in terms of the number and mix of students of various language groups and the language capacity of the school system staff, but also in terms of the goals of the community for those students. These goals are determined, if not always articulated, by the community, the parents, and the administrators, as well as by local, state, and federal policy makers and the educational staff. (p. 446)

The differences in program design are important for the different communities and when assessing the effectiveness of program. Thomas and Collier's (1997) longitudinal research on bilingual program outcomes propose that carefully scrutinizing for coherence of educational philosophy and program design, adequacy of implementation in terms of human, material, and financial resources, and consistency of their educational outcomes is critical. When these three factors are accounted for and there is an achievement gap reduction between children who are gaining proficiency in English and those who started school with native or native-like proficiency in English, then models could be compared in relation to their relative effectiveness. In addition, their research suggests that the effectiveness of bilingual programs is best measured longitudinally and with multiple tools. Bilingual students may initially need more time to perform at the levels suggested by national norms, but in the long term, and concomitant to a quality bilingual program, their gains are highly significant.

Performance assessment of bilingual students

While there are several competing perspectives as to what constitutes effective measures of student performance, ongoing assessment that includes providing teachers and students with feedback—for improving teaching and learning—goes unchallenged (Brisk, 1998; Brookover, 1985). Brisk (1998) argues that fair assessment of bilingual students requires three distinct sources of information: background knowledge of the students, understanding of the

processes students use to perform, and evaluation of the outcomes per se. This coincides with constructivist views of learning as a dynamic social process, as “an activity that is always situated in a cultural and historical context” (Bruner & Haste, 1987, p. 1). The role of the teacher in establishing a fair assessment of bilingual students’ developing skills cannot be overlooked, since they are the ones structuring the classroom experience through which bilingual children make sense of school activities by tapping and “translating” from the knowledge embedded in their linguistic and cultural background (Igoa, 1995).

Challenges in performance assessment data gathering

While few states do not have common statewide assessment in place, many school districts only report aggregated test results for either entire schools or district-wide programs, making the showcase of specific programs difficult (Charles A. Dana Center, 1999). In many districts, data are not disaggregated by program nor do they distinguish between student participants. The reporting of test results is usually an annual summative evaluation, in some cases to meet funding agency requirements (i.e., Title I or Title VII). Few school districts analyze their data longitudinally. Larson and Ovando (2001) suggest that aggregated test data reporting may be a way for school districts to purposely skirt issues of equity while using the rhetoric of equal opportunity. Whether it is the political will of school district leaders or the sheer difficulty in producing such detailed reports in times of personnel cutbacks, both disaggregation and the one-year reporting system affect bilingual education programs’ ability to showcase their success.

In many states bilingual students are not required to take standardized tests until they reach a minimum level of English proficiency, so their progress is neither measured with state assessment nor with native language assessment tools. Some bilingual education experts argue that exempting students from standardized testing practices is fair, since recent arrivals are not tested in content knowledge but rather in English proficiency (González, Castellano, Bauerle, & Duran, 1996). Furthermore, they argue that standardized tests are not culture-free. At best, test items may be representative only of the largest immigrant cultures in the country, leaving smaller language groups to take tests that are, from their perspective, culture biased. The counter argument launched is that, while the test results will not test knowledge per se, having them would serve to create a stronger sense of accountability in schools (Hakuta, 2001).

Brisk (2000) contends that a given program is successful by the attainment of its students, the amount of adverse factors it must conquer, and the quality of its practices. An imperative goal for bilingual research, hence, is to examine how different communities employ different measurable paths in the implementation of successful bilingual education. Thus, through our collective multicase study we attempt to bring together a comprehensive picture of what data are available and accessible in schools, what the data say about student performance, and how school staff conceptualize success, concomitant to the goals of PoS.

Methodology

The collective case study is a multi-site effort to inquire into a phenomenon in a variety of locations, with the expectation that their study will lead to a better understanding of similar sites (Stake, 1998). The accumulation of case study research promises a rich and robust picture of bilingual education, possibly benefiting practice and influencing policy (Cummins, 1999). Ethnographic techniques such as observations through shadowing, informal and formal interviews of the different participants, and document analysis were used, and standardized test scores were obtained for each of the schools. While we had specific ideas about the kind of data needed based on the PoS rubric (see www.lab.brown.edu/public/NABE/protraits.taf), we were open to examine information about assessment in bilingual/bicultural education programs more broadly, through rich descriptions and multiple data sources (Bernard, 2000; Denzin & Lincoln, 2000).

Settings, Demographics, and Participants

The three schools selected, one elementary and two middle schools, mirror the challenges found in many large urban school districts. We call them Burgos Elementary, Sol Middle, and Elliot Middle (see Table 1). Over 90% of the students in the three schools are eligible for free lunch. All three schools have long-standing bilingual programs; two are transitional programs, and one offers a dual language program. The schools have anywhere from 13% to 38% of their students classified as not proficient in English. All three programs have reputations for offering a quality education.

Table 1

School, Program, and Student Population Characteristics

School	Program Type	Grade Levels	Total Number of Students in School	Percentage of Students in Bilingual Program	Languages	Percentage of Students Receiving Free Lunch
Elliot Middle	transitional	6–8	530	13	Chinese/English	91
Burgos Elementary	transitional	pre-K–6	760	38	Spanish/English	95
Sol Middle	two-way	6–8	200	100	Spanish/English	90

Each one of the schools has a unique history. Elliot Middle boasts of having the first Chinese-English middle school transitional bilingual program in the school district, opening in 1971. The student body ethnic make-up in this school is composed of 32% African-American, 20% Caucasian, 22% Asian, and 25%

Hispanic students. The Chinese-English bilingual program is composed of approximately 70 students, and it has been showcased in local newspapers.

Burgos Elementary established its first bilingual program in 1971. It presently has a total population of 760 students from a diversity of ethnic groups: 76% Hispanic, 14% African American, 6% Caucasian, 3% Asian, and 1% Portuguese. The bilingual program in this school boasts of being fully integrated in the total school community, and the school had received an award for its service to the community.

Sol Middle is the first dual language middle school in the school district. It is designed as a small school, with no more than 200 students, mirroring the atmosphere of a private school. A school-watch parent organization wrote it up as among the best middle schools in the city. Ninety-eight percent of the students come from Spanish-speaking family backgrounds.

An average of 20 key informants for each site—teachers, administrators, counselors, students, parents, and community leaders—were initially identified through “community nomination” (Ladson-Billings, 1994, p. 147) for each school. The latter was important because we were aware of the potential information biases in gathering data from elite informants. The key informants represented a wide sample of the school community (Applebee, 1996). Our relationships as former professors of some of the school personnel required that we systematically analyze interview data and triangulate the information provided with additional data sources and peer debriefing procedures (Spall, 1998). We understood those triangulation opportunities not as a guarantee of validity, but rather as an alternative to it (Denzin & Lincoln, 1998).

Gaining access

The initial entry into the three schools presented different challenges. As we soon found out, “getting permission to conduct the study involves more than getting an official blessing” (Bogdan & Biklen, 1992, p. 82). There were issues of gatekeeping (Rossman & Rallis, 1998) that sometimes extend to personnel within the schools targeted for study. For two schools, proposals underwent the human subject review at the university and school district levels; the third school required state board of education permission. Once entry was negotiated, we requested written consent from all participants. In Elliot and Burgos, we met with principals and the teaching staff to discuss the proposal and the timeline of the study. Entry to Sol was facilitated because one of us has a long-standing academic relationship with the school. All relevant school personnel, students, and parents received an explanatory letter about the study in English and the native language—Chinese or Spanish—together with the consent form. Provisions for further clarification or to answer any questions were made.

The entry issues were associated with the accessibility of data. Trustworthy and ethical practices made possible the credibility of the project and helped us gain entrance to the insiders’ world, life, and thought of the participants (Baumann, 1996; Villenas, 1996). The school staff saw that our work in regards to PoS could benefit them (Rossman & Rallis, 1998, p. 43), that is, there was a

strong probability that their school would be showcased as having an exemplary program. Understanding the biases inherent in such a situation, hence, required that we take extra measures in maintaining accurate and detailed reporting of all methods and procedures employed (Janesick, 1998; Sarason, 1996).

Data sources and data analysis

The research questions in regards to what was available and its accessibility demanded that we use multiple data sources for an extended period of time. The data collection phase took between three to six months. We immersed ourselves in the schools' various settings (classrooms, hallways, library, teachers' lounge, principal's office, school auditorium, meetings, and school functions) and times, to learn the everyday regularities of the program (Rossman & Rallis, 1998). We shadowed students, interviewed teachers, administrators, parents, students, and staff associated with the bilingual program in and outside of the school building. We also took personal notes after informal contact with participants because that sort of communication proved useful to learn about potential sources of data and ways to access it.

Given that we had the PoS application (see <http://www.lab.brown.edu/public/NABE/nomform.shtml>), we decided to use it as the front-end data-gathering frame (Miles & Huberman, 1994). The use of the PoS instruments allowed a degree of structure to the study so that concepts were clarified and priorities for actual data collection were developed at the outset of the three case studies (Moll & Díaz, 1987). The interviews were developed as data sources to inquire into the participants' experiences in constructing evidence of success (Seidman, 1998).

We analyzed the data identifying trends and grouping participants by role for each school. Given the inordinate amounts of time spent in disaggregating data, we do not claim to have exhausted all the possibilities in the analysis of interviews. However, from the data analyzed, we gained a clear notion about the consistency of stances between school personnel and those served by them.

Limitations

Despite our attempt to ensure multiple data sources, there were issues related to accessibility that merit mention. The case studies were carried out with minimum funding and no release time for teachers. Based on their willingness to participate, we conducted interviews with teachers during their planning and development periods. The type and quality of information obtained through the teachers may have been limited by the contact-time and their perception regarding the importance of a research project carried out with a small budget and manpower. In addition, the schools studied were at the elementary and middle school levels. A wider sample including the high school level could have provided greater insights. In spite of those limitations, we feel confident that the wide net of informants, our follow up to clarify key

issues, the maintenance of field notes, and the length of time spent in the schools enabled us to arrive at redundancy of data gathering (Goetz & LeCompte, 1984).

Results

We report the findings that address our research questions in three distinct sections: data availability and accessibility, data in relation to student performance, and different perspectives that justified statements of success on the part of those associated with the programs studied.

Data Availability and Accessibility

One of our major findings relates to the wide range of significant data available to the schools/programs and to outsiders, presented as school district internal reports, local newspaper articles, and state and local reports and publications accessible online. Table 2 provides a summary of the data available in each of the three schools.

The data helped us to better understand those programs and the participants' perspectives in regards to them. Because the enormous amount of paper generated in and about schools is not always kept in any central location nor always accessible, gathering thorough and systematic documentation required creativity as much as persistence.

Table 2

Data Available and Source for Each School

	Data Available	Source
Elliot Middle	LAU Steps English proficiency (MELA-o & LAS) Native language proficiency (Chinese Cloze Test) Instructional levels in the the native language SPED designation (if any)	Bilingual specialist at district office LAU liaison
	Grades from report cards for all subject areas Data on students accepted into examination high schools	School counselor (former bilingual counselor for the school)
	State-mandated achievement test scores (MCAS) in mathematics, reading, and science (for a year prior to the study). For the years of the study, they became available six months after the test was administered.)	School district Web site (for the whole school. For bilingual students, they had to be specially requested by the bilingual program director)
	Standard 9 results for mathematics and reading (1999 and 2000)	Director of instruction
	Writing prompts (2 our of 3) Scholastic Reading Inventory (2 out of 3)	Turning Points Coach

Table 2 (cont.)

Data Available and Source for Each School

	Data Available	Source
Burgos Elementary	<p>LAU Steps and English proficiency Native language proficiency Instructional level in the native language</p> <p>State-mandated achievement test scores in English: mathematics, reading, and writing (for previous and current year of the study)</p> <p>District-mandated achievement test scores in native language: reading, mathematics, and language. SFA English and Spanish reading tests and APRENDA (reading, language arts, and mathematics in Spanish)</p> <p>District scores for school reform program</p>	<p>School principal</p> <p>State department of education report for the school and website (for the whole school. For bilingual students, they had to be manually accomplished using school rosters)</p> <p>Districts Office of Research and Evaluation</p> <p>School principal</p>
Sol Middle	<p>LAB scores in English LAB scores in the native language Special Education Status Recent immigration/Place of birth Attendance Language use</p> <p>State and city mandated test scores in English (CTB-Reading; English Language Arts Achievement; in mathematics (CTB-Mathematics and PAM; Mathematics Achievement; Spanish-reading (El Examen de Lectura en Español-ELE); writing test (PAL); science test; technology, and social studies Spanish and Math Regents exam Writing samples</p> <p>Extracurricular activities</p> <p>Program/curricular/language use descriptions</p> <p>Parent/community/support</p>	<p>Annual school report Comprehensive Education Plan 2000–2001 Acting principal Students</p> <p>Annual school report Comprehensive Education Plan 2000–2001 Acting principal</p> <p>Annual school report Comprehensive Education Plan 2000–2001 Acting principal Observation Publications</p> <p>Annual school report Comprehensive Education Plan 2000–2001 Acting principal</p> <p>Parent/teacher Interviews</p>

Often times the data were presented in a manner that required reorganization so that it would become useful for our project. For instance, in one of the schools we found it difficult to summarize and interpret the state mastery tests in reading and math, because not all the bilingual students had taken the test and the report compared average scores for different groups of

students. We could only surmise that the increase in student mobility and/or the exemptions of bilingual students from testing might help explain the discrepancy. In all cases, making data reader-friendly required many hours and specific know-how. Obtaining data on bilingual student performance sometimes required going to the raw scores with a master list of students participating in the bilingual program and extracting the information so that it could be reanalyzed. In Elliot, we decided to highlight the processes of gathering data on bilingual eighth graders because of the existence of an additional state-mandated standardized test. We also assumed that the success of the program could be best measured by concentrating on the students with the longest stay in it. The data available for Elliot's eighth graders are summarized in Table 3. The data were not available in one place and to illustrate how resourceful we needed to become in the process of data gathering, we include information on who provided the data for each measure.

For example, the data needed in order to establish an English instructional placement level, determined by the Lau Step categories, are in columns 1 through 5. The data (date of entry, Lau Step, literacy and instructional levels in Chinese and English, and SPED designations if any) were obtained for 26 out of 27 bilingual students. While the English proficiency and English literacy tests, MELA-O and LAS, are used to establish Lau categories (column 4), at Elliot it was a committee composed of bilingual teachers and the parents of the target student that established the ranking of the student, based on proficiency and literacy test scores and other information. Items 3 and 3a provide information on literacy levels in the native language and English. Item 3 is measured by the Chinese Cloze Test. The test scores indicate students' Chinese proficiency as of June 1999. The categories with respect to native language literacy levels are NNP, LNP1, LNP2, and FNP (non, limited, and fluent native proficiency). Column 3a provides the instructional placement level in the native language for 20 of the bilingual eighth graders as of January 2000. The number represents the instructional level in relation to grades 1-12 of native language performance and is based on the Chinese Cloze Test. The Lau Liaison—a former bilingual teacher who worked in the district's central office but serviced schools in the area—provided the information in columns 1 through 6.

The student outcome data available to us (columns 6 through 12) were Stanford 9 (S9), Writing Prompts (WP), and Scholastic Reading Inventory (SRI) scores. These type of data would allow programs to showcase success in comparison to district and statewide norms. The S9 is a well-known norm-referenced test that measures student achievement in reading and math. The

Table 3

Sample of Available Data at Elliott Middle (Abbate & Brisk, 2001)

1- Entry Date	2- LauStep	3- Native Lang. Profi- ciency	3a- Instr. Level in NL	4-Eng. Profi- ciency	5-SP- EDD- es.	6-S9- 199- 9Re- ading	7-S9- 2000- Readi- ng	8- S9- 2000- Math	9- WP Oct. 99	10- WP- Feb. 00	11- SRI- Oct. 99	12- SRI- Feb- 00	13-ESL			14- Reading		15-Math	
													L	F	Fx	F	Fx	F	Fx
3-97	1		8.5	LEP1		17	17	57					2	B	B-	A-	B	B-	
8-98	1		8.5	LEP1	R2	4	4	30					1	C-	C		A-	A-	
5-99	1	LNPI	8.5	NEP				66					1	F	F	D-	D	F	
5-00													1	D-	D+	B	A	B-	
12-95	2	LNP2	8.5	LEP2		15	15	62	1	1.5	730	645	4	A-	C	B	B	C+	
10-99	1	LNP2	8.5	NEP				81					1	B	B	B+	A-	B-	
4-99	2	LNPI	8.5	LEP2				91	1.75		645	675	3	A	A-	B	B	A	
7-99	2	LNP2	8.5	LEP2				94	2			950	1	A	A	A	B	A-	
8-99	1	LNP2	8.5	NEP				26		2			1	F	F	D-	D-	F	
7-90	4	LNPI	0	TEP	R3	46	46	83		2			M	B	B+	C+	D	B	
9-95	2	LNP2	5.0	LEP2		14	14	64	1.5		715	690	3	A-	B-	B-	C	D+	
9-99	1	LNP2	8.5	NEP				34					1	D	F	D	C	C-	

data provided in Table 3 are for a two-year period (May 1999 and May 2000), and the number of students taking the exam depended on their Lau classification. In May 1999, for example, scores for 10 out of the 27 students who were then enrolled in seventh grade were reported. The remaining 17 students were not tested. Twelve students were arrivals from September 1999 or later, and five had entered the school a month or less before the test was administered. By May 2000, only one of the 27 students was exempted from taking the S9, based on an internal decision by the school's director of instruction that all students, regardless of status, be tested.

We obtained the results of the May 2000 results of the S9 because the director of instruction was exhilarated about the relative gains in math and reading for the entire school upon receipt of the data. He bragged about student learning at Elliot and proudly provided copies requesting that confidentiality be maintained. The WP and SRI scores, although administered by the school, were organized and kept by the Turning Points coach, who visited the school approximately once a week. We obtained the data in columns 9, 10, 11, and 12 through e-mail. The coach could not, however, answer our questions in regards to why four or five students out of potentially 10 who were Lau step 2 or above had been tested, neither who was responsible for testing those students with such tools. We could not obtain WP and SRI scores for the final testing session of June 2000, because apparently the coach never got a hold of the scores. This same level of difficulty in data gathering was not necessary in all the schools studied for Sol had standardized data that was disaggregated by the district central office as a matter of policy.

Elliot and Burgos teachers alike spoke about their assessment with teacher-made tests during interviews but did not make them available, whereas the Sol Middle teachers provided us with multiple classroom-based assessment tools. When we shadowed students, we witnessed that teachers taught well-prepared lessons and that students were engaged in instructional activities. In Elliot and Burgos, much of the daily lesson assessment was done through informal observation of the students (Stefanakis, 1998). At Sol Middle the teachers formally met to assess all students individually.

Data in Relation to Student Performance

To answer the question of data in relation to student performance posed in our study, we start by further analyzing the Elliot eighth graders' scores. The S9 math scores were significant (Table 3, column 8). Seventy percent of the students scored above the 50th percentile and 33% above the 75th percentile, an outstanding feat considering the short time many of the students had been in American schools. The S9 reading scores, on the other hand, were low (column 7). Almost 89% of the students scored below the 25th percentile. The S9 reading, WP (columns 9 and 10), and SRI (11 and 12) scores were consistently low. As Table 3 shows, WP and SRI were not administered consistently to bilingual students mainstreamed in reading (10

students in categories Lau Step 2 and above). Both were assessment tools mandated by the school district to hold teachers accountable for teaching basic literacy skills—and to increase the district’s low standing with respect to student reading achievement in statewide testing.

Another significant finding was the existence of few measures of student progress in the native language. Native language literacy was assessed in all three schools, although through different means—Burgos and Elliot used cloze tests and Sol used a citywide test. Burgos and Elliot, being transitional programs, did not place a big emphasis on native language maintenance. Sol belonged to a school district where some content area knowledge was measured in the students’ native language, but students had to be eligible to take these tests. The State Regents exam in Spanish was most significant with respect to how the Spanish teacher connected the test to instruction. Students were placed in an advanced Spanish class, and the teacher assessed each of her students to determine their eligibility to take the Spanish language exam, which was tied to high school requirements.

At Burgos, student outcomes in reading were heavily related to the school’s adoption of the Success for All (SFA) school-restructuring plan. Through heterogeneous groupings of readers (in both Spanish and English), the number of students reading at or above grade level had increased considerably. Scores from tests administered every eight weeks were available and allowed teachers to move students accordingly.

Perspectives on Program Success

Program directors associated with all schools were likely to spout out what percent of their students had tested above the 50th percentile or how many had passed demanding entrance exams for further education the previous year but could not provide disaggregated data that supported such statements. At Elliot, the mainstreaming patterns became self-revealing as to how Elliot’s staff thought about their success. The bilingual program did not follow traditional partial-mainstreaming patterns (see Brisk, 1998) and kept bilingual students the longest possible time in bilingual math and science. Past bilingual student success in math and science statewide tests, their acceptance rate at the math-and-science centered examination high school, and in district wide competitions and fairs, supported the bilingual team’s implementation of a de facto advanced placement program for math and science. The school administration gave its tacit approval to the practice.

The trend of Asian origin students’ higher performance in math than in verbal skills has been previously documented (Stodolsky & Lesser, 1967). The ways in which the school community explained the gap between math and reading test scores were identified in interviews with parents and students. First, teachers pointed out that the math curriculum in China’s sending schools seemed to demand more from students as compared to the grade level requirements in

American schools. Bilingual students felt they had a head start when they changed school systems. Second, some parents reported helping their children with math problems and homework. Math was perceived as the least language-based content area, and parents felt they could participate in helping their children learn. Some parents reported with pride that they added more math homework to what their children brought home from school, which mirrors prior research findings by Hess, McDevitt, and Chih-Mei (1987) on Chinese parents' attitudes about schoolwork and math success. Thus, student attainment was associated with the curriculum of the sending country and parental support. In addition, language was not viewed as a significant barrier to success. Measures such as the WP, however, were perceived to be at odds with the native culture. Within the Elliot Chinese community culture, talking about self or expressing opinions is not likely to be viewed as appropriate behavior for children. Parents reported that in Chinese schools students were assessed in their writing based on the knowledge of facts, proper grammar, and vocabulary. Writing assessments never had to do with feelings elicited by a certain story or event, and fiction was not highly regarded as a vehicle for acquiring academic content. Thus, the bilingual staff's arguments regarding the WP were critical to understand how they explained the gap in assessment results.

On the other hand, success at Burgos was related to mainstreaming and the impact of SFA. Teachers proudly showed that their students were moving up at least one Lau step per year. Those students who showed rapid signs of progress were mainstreamed in one or more classes. However, as teachers pointed out, it was often the student with a strong command of Spanish literacy who had little trouble in becoming fluent in their second language. The quality and low mobility of the bilingual staff were also mentioned as essential to the school's success. The competitiveness imbedded in SFA also seemed to play a role. The SFA coach showcased those classrooms with the highest number of students passing the test, and scoring at or above grade level, thus tacitly imposing an in-house benchmark for responsible teachers to approximate. Specific staff development tailored to teachers' needs was also in place.

At Sol, teachers presented a complete picture of their daily work and the ways it led to success. They provided us with samples of teacher-made tests, samples of student writing in content area, and language state tests from the creative writing class. During the interviews, the teachers elaborated on the ways they assessed their children. Sol's teachers talked about taking into consideration, for both instruction and assessment, the students' comfort level in the language, their personalities, their literacy levels, their basic skills levels and their sustained ability to function academically. They accounted for adjustment and adaptation of instruction for the diverse needs of their students (Taylor & Nolen, 1996). They created their own quizzes in the content areas because of their collective lack of confidence in the assessment instruments provided with commercial curriculum guides to measure what was taught.

Teachers were in agreement that one of the strongest indicators of how well students performed in tests was related to their strengths in writing. They pained over the fact that low writing achievement scores did not do justice to the successful writing products completed by the same students in class.

Discussion

The collective case study revealed that there are significant data available and accessible to the schools/programs in the form of internal reports, newspaper articles, publications, and state and local reports publicly accessible. The data were rarely disaggregated in ways that enabled the schools to highlight the bilingual program. While the issue of resistance to disaggregating the data (Larson & Ovando, 2001) might be applicable to school districts in general, this explanation did not help clarify what happened in the three cases we studied. In these schools, the lack of data availability and access was not related to fears of embarrassing student outcomes directly, for the opposite was true. The bilingual students did very well when compared to their mainstream counterparts. The schools' failure to present the data required by PoS was related to a lack of human and financial resources and know-how. Our research confirmed that those schools needed help in gathering and analyzing the evidence required for showcasing the achievement and potential of their programs in PoS.

A problematic research finding is related to who gets tested. Students who are not proficient in English, however they are defined locally, are often exempt from district and state testing. The assumption is that most of the bilingual students who do not take the exam have not mastered English sufficiently to ensure that they are being tested about anything other than English, and that standardized tests reveal little about developing bilingual students' potential (Gonzalez et al., 1996). In two out of the three cases, only recent arrivals were exempt; in the other, students were exempt for the first three years. Testing policies for bilingual students were not held consistently either, as we saw in Elliot, where suddenly all recent arrivals were required to take standardized tests by the end of the year, a move perhaps motivated by issues of accountability (Hakuta, 2001).

The use of standardized assessment systems for monitoring and improving teaching and learning is still an area in development (Brisk, 1998; Gipps, 1999). Within the context of constructivist learning, standardized tests alone do not show overall student relative gains in the way that portfolios or interactive assessments do. Even for the limited areas of need that can be inferred from analyzing standardized test results, there was little evidence that tests scores were used for anything other than the exhibition of progress of the entire student body in all three schools. Teacher involvement in the establishment, review, and monitoring of assessment systems was also minimal.

Even when there was evidence of involvement in some collective assessment of students, as in Sol's and Elliot's case, the voices of these teachers at the district, state, or national level in establishing assessment systems were almost non-existent, as it is for most public school teachers.

Defining and measuring what constitutes success in education for the public view has virtually become the domain of standardized test-manufacturers and eager politicians. Those with an emic perspective in the education and assessment of bilingual students—teachers and researchers alike—need to assert their thorough understanding of relevant issues in ways that are accessible to the wider public.

Despite the challenges reported, we found that the ways schools constructed their views of self as successful were substantiated. Student performance ranged from average to high depending on school subjects. Teaching practices agreed with what the communities in which the schools were immersed expected from their teachers (Brisk, 2000; Moran & Hakuta, 1995). The starting point for many bilingual programs does not compare to the regular education programs in that the level of native language literacy or even prior schooling differs significantly. Thus, the expectations for bilingual programs surpass those of regular education programs. Bilingual programs are expected to do more for students who start behind in order to catch up and be treated and judged the same as other children. The rich descriptions of the context, resources, processes, and outcomes in which instruction and assessment take place in the three schools helped us understand how schools used their own notions of success and how student achievement data were part of those notions. Schools tend to use achievement test results as a general measure of how rapidly they prepare the students for participation in an all-English curriculum. Two of the schools turned to encouraging math scores to illustrate the strength of the social equalizing role of their schools. They claimed they were capable of doing more for those who start behind, that there was value added to the education of children as a result of their participation in bilingual programs. While we gathered data that showed those schools did merit their reputations, we also found additional ways in which the data could be used to improve teaching and learning.

More importantly, we assessed that school district evaluation systems were neither user-friendly, nor made available as a matter of public record. Teachers also often delegated the collection and compilation of formative assessment to the school administrators and thus, did not have the information they needed readily available to improve teaching and learning. Since only a handful of the teachers interviewed provided actual evidence of their own classroom assessment strategies, this persuaded us of their reliance on observation for monitoring students (Stefanakis, 1998).

Conclusion

Our study strongly suggests that more staff development for teachers who work with bilingual students is necessary to address the void in classroom assessment data, and to equip practitioners to understand student assessment techniques in the broader context (Olebe, 1999). As the National Commission on Teaching America's Future (1996) states in its report, *What Matters Most*, student academic achievement is bound to increase in the presence of high quality teachers who use assessment to monitor students' learning and their own teaching. Furthermore, it is our belief that if this does not occur, schools will not build on their ability to showcase the added value of thoroughly designed programs, their efforts toward equity in the face of increasing student diversity, and the social equalizing nature of their bilingual programs.

References

- Abbate, J., & Brisk, M. E. (2001, April 10–14). *Measuring success in an urban Chinese transitional bilingual program*. Paper presented at the Annual Meeting of the American Educational Research Association, Seattle, WA.
- Applebee, A. N. (1996). *Curriculum as conversation: Transforming traditions of teaching and learning*. Chicago: University of Chicago Press.
- Baumann, J. F. (1996). Conflict or comparability in classroom inquiry: One teacher's struggle to balance teaching and research. *Educational Researcher* 25 (7), 29–36.
- Bernard, H. R. (2000). *Social research methods: Qualitative and quantitative approaches*. Lanhan, MD: Rowman & Littlefield Publications Group.
- Bogdan, R. C., & Biklen, S. K. (1992). *Qualitative research for education: An introduction to theory and methods* (2nd ed.). Boston: Allyn & Bacon.
- Bran, D. (1987, August 31). The new whiz kids. *Time*, 130 (9), 42–51.
- Brisk, M. E. (1998). *Bilingual education: From compensatory to quality schooling*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Brisk, M. E. (2000). *Quality bilingual education: Defining success* (Working Paper No.1). Providence, RI: Brown University, Northeast and Islands Regional Educational Laboratory.
- Brookover, W. B. (1985). Can we make schools effective for minority students? *The Journal of Negro Education*, 54 (3), 257–268.
- Bruner, J., & Haste, H. (1987). *Making sense: The child's construction of the world*. New York: Routledge.

- Charles A. Dana Center (1999). *Hope for urban education: A study of nine high-performing, high-poverty, urban elementary schools*. Washington, DC: U. S. Department of Education, Planning and Evaluation Service.
- Cummins, J. (1999). Alternative paradigms in bilingual education research: Does theory have a place? *Educational Researcher* 28 (7), 26–32, 41.
- Darling-Hammond, L. (1998). Teacher learning that supports student learning. *Educational Leadership* 55 (5), 6–11.
- Denzin, N. K., & Lincoln, Y. S. (1998). Entering the field of qualitative research. [Introduction]. In N. K. Denzin & Y. S. Lincoln (Eds.), *Strategies of qualitative inquiry* (pp. 1–34). Thousand Oaks, CA: Sage Publications.
- Denzin, N. K., & Lincoln, Y. S. (Eds.). (2000). *Handbook of qualitative research* (2nd ed.). Thousand Oaks, CA: Sage Publications.
- Gipps, C. (1999) Socio-cultural aspects of assessment. In A. Iran-Nejad & P. D. Pearson (Eds.), *Review of research in education* (Vol. 24, pp. 355–392). Washington, DC: American Educational Research Association.
- Goetz, J. P., & LeCompte, M. D. (1984). *Ethnography and qualitative design in educational research*. New York: Academic Press.
- Gonzalez, V., Castellano, J. A., Bauerle, P., & Duran, R. (1996). Attitudes and behaviors toward testing-the-limits when assessing LEP students: Results of a NABE-sponsored national survey. *Bilingual Research Journal* 20 (3/4), 433–463.
- Greene, J. P. (1997). Meta-analysis of the Rossell & Baker review of bilingual education research. *Bilingual Research Journal* 21 (2/3). [On-line]. Retrieved February 5, 2002, from <http://brj.asu.edu/archives/23v21/articles/art1.html>.
- Hakuta, K. (2001, April). *The education of language minority students: Testimony to the U.S. Commission of Civil Rights*. Retrieved February 5, 2002, from <http://www.stanford.edu/~hakuta/Docs/CivilRightsCommission.htm>
- Hess, R., McDevitt, T., & Chih-Mei, C. (1987). Cultural variations in family beliefs about children's performance in mathematics: Comparison among People's Republic of China, Chinese-American, and Caucasian American families. *Journal of Educational Psychology* 79, 179–188.
- Igoa, C. (1995). *The inner world of the immigrant child*. New York: St. Martins Press
- Janesick, V. J. (1998). The dance of qualitative research design: Metaphor, methodolatry, and meaning. In N. K. Denzin & Y. S. Lincoln (Eds.), *Strategies of qualitative inquiry* (pp. 35–55). Thousand Oaks, CA: Sage Publications.
- Ladson-Billings, G. (1994). *The dreamkeepers: Successful teachers of African American children*. San Francisco: Jossey-Bass Publishers.

- Larson, C. L., & Ovando, C. (2001). *The color of bureaucracy: The politics of equity in multicultural school communities*. Belmont, CA: Wadsworth/Thomson Learning.
- Miles, M. B., & Huberman, A. M. (1994). *An expanded sourcebook: Qualitative data analysis*. Thousand Oaks, CA: Sage Publications.
- Moll, L. C., & Diaz, S. (1987). Change as the goal of educational research. *Anthropology and Education Quarterly*, 19, 300–311.
- Moran, C. E., & Hakuta, K. (1995). Bilingual education: Broadening research perspectives. In J. A. Banks & C. A. M. Banks (Eds.), *Handbook of research on multicultural education* (pp. 445–462). New York: Simon & Schuster Macmillan.
- National Commission on Teaching and America's Future. (1996). *What matters most: Teaching for America's future*. New York: NCTAF. Retrieved February 5, 2002, from <http://www.nctaf.org/publications/WhatMattersMost.pdf>
- Olebe, M. (1999). California formative assessment and support system for teachers (CFASST): Investing in teachers' professional development. *Teaching and Change* 6 (3), 258–271.
- 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. Final report to the U. S. Department of Education* (Executive Summary and Vols. 1 & 2). San Mateo, CA: Aguirre International.
- Rossmann, G. B., & Rallis, S. F. (1998). *Learning in the field: An introduction to qualitative research*. Thousand Oaks, CA: Sage Publications.
- Sarason, S. (1996). *Revisiting the culture of school and the problem of change*. New York: Teachers' College Press.
- Seidman, I. (1998). *Interviewing as qualitative research: A guide for researchers in education and the social sciences*. New York: Teachers College Press.
- Spall, S. (1998). Peer debriefing in qualitative research: Emerging operational models. *Qualitative Inquiry* 4 (2), 280–292.
- Stake, R. E. (1998). Case studies. In N. K. Denzin & Y. S. Lincoln (Eds.), *Strategies of qualitative inquiry* (pp. 86–109). Thousand Oaks, CA: Sage Publications.
- Stefanakis, E. (1998). *Whose judgement counts? Assessing bilingual children, K–3*. Portsmouth, NH: Heinemann.
- Stodolsky, S., & Lesser, G. (1967). Learning patterns in the disadvantaged. *Harvard Educational Review* 37, 546–593.

- Suarez-Orozco, M., & Suarez-Orozco, C. (2000). Some conceptual considerations in the interdisciplinary study of immigrant children. In E. Trueba & L. Bartolome (Eds.), *Immigrant voices: In search for educational equity* (pp. 17–35). Boulder, CO: Rowman & Littlefield Publishers, Inc.
- Taylor, C. S., & Nolen, S. B. (1996). What does the psychometrician's classroom look like? Reframing assessment concepts in the context of learning. *Educational Policy Analysis Archives* 4 (11). Retrieved February 5, 2002, from <http://epaa.asu.edu/epaa/v4n17.html>
- Thomas, W. P., & Collier, V. P. (1997). *School effectiveness for language minority students*. Washington, DC: National Clearinghouse for Bilingual Education.
- Tse, L. (2001). *Why don't they learn English? Separating fact from fallacy in the U.S. language debate*. New York: Teachers College Press.
- Villenas, S. (1996). The colonizer/colonized Chicana ethnographer: Identity, marginalization, and co-optation in the field. *Harvard Educational Review* 66 (4), 711–731.
- Willig, A. (1985). A meta-analysis of selected studies on the effectiveness of bilingual education. *Review of Educational Research* 55, 269–317.
- Willig, A. C. (1987). Examining bilingual education research through meta-analysis and narrative review: A response to Baker. *Review of Educational Research* 57, 363-76.

Endnotes

¹The Northeast and Islands Regional Educational Laboratory at Brown University sponsored the case studies.

²A joint project of the National Association for Bilingual Education, the Northeast and Islands Regional Educational Laboratory at Brown University, and Boston College. The website may be found at www.lab.brown.edu/public/NABE/protraits.taf

³All names are pseudonyms.

⁴Lau Step 1 indicates an absence of English proficiency; students are enrolled in bilingual classes one hundred percent of the time. When they reach Step 2, students are mainstreamed in reading. In Step 3, social studies and language arts are added. In Step 4, students are fully mainstreamed but monitored by the bilingual staff for one year.

⁵The 27th student had just arrived and had not been tested.

⁶The Massachusetts Comprehensive Assessment System (MCAS) data for the 2000 cohort of eighth graders were unavailable at the end of the data collection phase. We did obtain MCAS data for the previous year eighth graders from the department in charge of data collection and analysis in central office. That department did not respond to our query or call put through by Elliot's principal. Eventually, the data were released with the help of the bilingual program director who oversees all bilingual programs in the school district. The 1999 MCAS data were of little help because for our study because it pertained to other students. However, those MCAS results were showcased in many ways by school staff.

⁷The principal had recently been awarded recognition for turning the school around from a widely known private foundation.

⁸The maximum WP score was 4.

⁹Four of the six bilingual students tested scored below 800 along with 15% of all eighth graders. One scored between 901 and 1000 along with 19% of all eighth graders, and the remaining one scored between 1001 and above, along with 51% of all eighth graders tested in the school.