The Acquisition of Academic English by Children in Two-Way Programs: What does the Research Say?

Stephen Krashen
Presented at NABE, 2004, Albuquerque, NM

Two-way bilingual programs are those in which majority language children and minority language children attend the same classes at the same time, the idea being that majority language children will acquire the minority language and minority language children will acquire the majority language. There has been a great deal of excitement about these programs; claims have been made that two- way has been shown to be the most effective form of bilingual education, and the best possible program for language minority children in general. It is important that we take a careful look at the research to see if these claims are supported.

The purpose of this paper is to examine one aspect of the research on two-way programs, specifically to examine the progress of language minority students in acquiring academic English. The focus here is on performance on tests of reading comprehension, a measure considered to be the gold standard for academic English. The questions to be discussed are these: (1) Do English learners in two-way programs show evidence of significant acquisition of English? (2) Do they outperform children in non-bilingual (all-English) options? (3) Do they outperform children participating in other forms of bilingual education? Some data is also available comparing two-way children with native speakers of English.

I will review the studies in ascending order of their ability to answer these three questions. First, I briefly discuss studies that had to be excluded from this analysis. Then, we look at studies that lacked comparison groups. These studies are able to provide us with some information regarding the first question. The third group consists of studies in which two-way students are compared to similar students in mainstream or submersion classes, an attempt to answer question (2). The fourth group consists of studies comparing two-way students to those in other kinds of bilingual education programs, an attempt to answer question (3). Finally, I include a section on studies in which two-way children are compared to native speakers of English.

Group One: Excluded Studies

I begin this review by excluding a number of studies from the analysis, reports in which few details are reported (Thomas and Collier, 1997), reports in which it is not

possible to determine how well English learners did because their scores are combined with those of native speakers of English (Christian, Montone, Lindholm, and Carraza, 1997; Rhodes, Crandell, and Christian, 1990, Lindholm and Fairchild, 1990, Sensec, 2002) .and reports in which two-way students were not tested in English (Coy and Litherland, 2000). In addition, in several of these studies, a substantial percentage of minority language speakers appear to have already acquired a considerable amount of English before beginning the program. These studies do not shed light on any of the three issues to be discussed.[1] Two studies labeled "dual language" comparisons were excluded because comparisons were not monolingual speakers of English but were bilingual but English-dominant (Perez and Flores, 2002; Oller & Eilers, 2002).

Group Two: Studies without Comparison Groups

Sugarman and Howard (2002) studied eleven two-way programs, and reported that Spanish speakers in these programs improved on different tests of English from grade three to five. There was, however, no comparison group and the measures used were not standardized. Thus, there was no way to compare the two-way childrens' progress to children in other programs. Sugarman and Howard's results, however, provide a positive answer to the first of the three questions posed in the introduction; they show that children in two-way programs make progress in acquiring English.

Several studies provide data using standardized tests without comparison groups. Before proceeding with a discussion of these studies, it should be noted that results in these studies are reported as NCE (Normal Curve Equivalent) scores. NCE scores are very useful when comparing across studies: they are constructed so that the mean is 50 and the standard deviation is 21.06.

NCE scores are not the same as ordinary percentiles, except at 1, 50, and 99. Of special interest is the fact that when scores are below the 50th percentile, NCE scores are higher than percentiles: A percentile rank of 29, for example, is equalivant to an NCE score of 38. Above the 50th percentile, NCE scores are lower than percentiles. A percentile rank of 76 is equivalent to an NCE score of about 65. Thus, in discussing the studies to follow, both NCE and percentiles will be included. (I thank Jim Crawford for pointing out the necessity of doing this.)

de Jong (2002) reported that different cohorts of Spanish-dominant children in a two-way program consistently attained NCE (Normal Curve Equivalent) scores between 38 (29th percentile) and 43 (37th percentile) on the Stanford Reading Test in grade

five between 1995 and 2000. Sample sizes were small, however, ranging from 13 to 18 students. A possible problem with this study is that “Spanish-dominant” does not necessarily mean limited English proficient; no information was provided on subjects’ initial English proficiency.

In Lindholm (2002), on tests of English reading, English learners in three two- way programs achieved NCE scores of 34 (23rd percentile), 35 (24th percentile) and 39 (30th percentile) by grade six, with one group achieving an NCE score of 43 (37th percentile) by grade seven (see table 10.6, p. 217). (One group, however, attained only the 9th percentile in grade five, see discussion below.) A subgroup studied longitudinally (n = 149) attained an NCE score of 41 in grade seven (34th percentile). Because few subjects were classified as orally proficient in English in kindergarten (Lindholm, 2002; p. 190) we can be confident that most subjects in this group were in fact limited in English when they began the program (see footnote 1 for studies in which this was not the case).

Thomas and Collier (2002) reported that a larger percentage of Spanish-speaking children at a two-way program at the Grant School in Oregon exceeded state norms than did students state and district-wide. This is impressive, but it was only true for one of the two groups tested, consisting of 12 students in one grade (grade 3). In grade 5, Grant students do not do well, scoring lower than district and state norms, but Thomas and Collier note that the Grant group (only 17 children) included late-comers. The Grant school has a very high mobility rate (71% annual). There was no separate analysis of those who have been in the program since kindergarten.

A multiple regression analysis, however, revealed a positive and statistically significant relationship between years in the program and scores on the English reading test, controlling for SES (poverty). Moreover, the effect of years in the program was larger than the effect of SES.

It is difficult to determine whether children at Grant began kindergarten with no English. It is unlikely. At grade 1, Grant students in scored 13.81 on the SOLOM measure of oral English, a test with a range of 5 to 25; a score of 19 is considered “proficient.” By way of comparison, English speakers in the two-way program did not reach this level in Spanish until grade 4.

Group Three: Comparison to English Learners in the Mainstream

We would expect English learners in properly organized two-way programs to outperform comparison children in the mainstream, especially those in

“submersion” programs.

In Ajuira (1994), "Hispanic" two-way students in grade 1 did better than comparison students in a mainstream class on a modified version of the Iowa Test of Basic Skills. Ajuria modified the test herself, making it shorter, including only items of “average difficulty” and items rated “fairly easy” (p. 139). The scores ranged from zero to six, but Ajuira does not tell us how this scale was constructed.

The two-way students scored 3.7 on a 0-6 scale and the mainstream students scored 2.4. Native speakers in the mainstream scored 3.4, below the level of the two-way Hispanic children. Before calling this a victory for two-way, however, there are problems:

1. Two-way students were given the same test the previous semester and scores were identical. There were, therefore, no gains.
2. Only 11 students were tested in the two-way program and only seven in the mainstream program.
3. We do not know if the Hispanic students were ever limited in English.
4. The mainstream classroom was an inferior pedagogical environment: “Materials were painfully lacking in the mainstream classroom, to the extent that there was only one Big Book in the entire room, and the other few books there were occupied less than a quarter of one bookshelf. The readers the students used in the small reading groups were kept by the teacher, and were only available to the students during reading time ... In both classrooms of the bilingual program, in contrast, there were well over twenty Big Books in Spanish and in English ...” (p. 88). Native speakers of English in the two-way program did much better than those in the mainstream program (4.4 versus 3.4). This could be because of the value of the two-way program or the impoverished print environment of the mainstream classroom (or both). Perhaps the correct generalization is that two-way students did better than those who were in submersion in a print-deprived environment.

Castillo (2001) reported that “native Spanish speakers” ages 5 to 8 (K-2) in two-way did far better than comparisons in a “regular classroom” (41st vs. 11th percentile on ITBS Reading), but only four children were in the comparison group. We do not know if the “native Spanish speakers” were limited in English (although it appears that those in the submersion class were). Results were also quite variable: the standard deviation for the two-way group was 26.

The two-way students also did better than the full group of comparison students (mean = 34, sd = 27); 30 out of 34 of the comparison children were native speakers

of English.

Castillo examined the impact of years in the two-way program, reporting very high scores the first year, and lower scores for those in the program longer. The very high score after grade 1 is unusual, and suggests that not all the children were limited in English.

Table 1
Years in Program and ITBS Scores (percentiles)

years in program	n	mean
1	21	57
2	20	26
3	20	34

from: Castillo (2001)

The results of this study are difficult to interpret. To be sure, two-way children did better than comparisons, both Spanish and English native speakers. But there were only four children in the Spanish-dominant comparison group, and scores were very high at the beginning but lower for those in the program two and three years. The data could mean that two-way is highly successful, and it could mean the opposite, that children begin the program with high levels of English proficiency and then get worse. In addition, the lower scores could also be a result of high-scoring children exiting the program.

Collier and Thomas (2002) provide some additional data. Two-way children in Houston appeared to be well ahead of children acquiring English whose families refused special help for their children. These “refusers” scored at the 20th percentile level at grade five (NCE = 32). Children in ESL-only programs, however, reached the 66th percentile in grade 5 (NCE = 59), doing better than two-way children. Two-way children in Houston scored at the 52nd percentile in grade 5 (NCE = 541.3). It is possible, however, that those assigned to ESL had higher levels of English when entering school, and the data are not controlled for poverty and neighborhood, as are students in Collier and Thomas’ other Houston comparisons (see below).

Group Three: Comparison with Other Bilingual Education Programs

Comparison to Transitional Bilingual Education

In Clayton (1993), “Spanish-speaking” students who had participated in transitional bilingual education outperformed “Spanish-speaking” students who had been in a two-way program on English reading. The two-way students had been in the program for at least four years. The results were similar when parental education was controlled, but the difference was greater when parents in both groups had no education. Two-way students outperformed Native American students from similar SES status. Students were in grades 3 through 8. We do not know if “Spanish-speaking” meant limited in English, however. Clayton suggests that the two-way program had an inferior reading program (p. 155) and was a newer program (p. 157). Table 2 presents data from two successive years of testing.

Table 2:
Comparison of Two-way and TBE Graduates, Grades 3-8

	1991	1992
	n/mean	n/mean
2-way	83/662.23	83/690.41
TBE	59/686.61	59/708.8

From: Clayton(1993)

Cazabon, Lambert and Hall (1993) report on the English proficiency of Spanish-speaking children in a two way program in Massachusetts, as compared to children “in a standard bilingual program” (p. 5). Two-way and comparison children were similar in social class and intellectual ability, as measured by Raven Matrices test.

It is not clear that all children classified as Spanish speakers (“Spanish amigos”) were Spanish dominant. In grade 1, for two cohorts combined, seven out of 47 were considered English-dominant (Lau classification score of 4 or 5 out of 5), and none were classified as completely Spanish-dominant. This measure was not applied at the beginning of the program, but seven months into the first grade; it is quite possible that the children had acquired considerable English conversational competence by then.

Table 3 presents scores on the California Achievement Test for two cohorts. The results are impressive: two-way children outperform comparisons in all grade levels

for both cohorts, and achieve above the expected (grade level) score in grade 3 in one cohort. There are, however, reasons to be somewhat cautious:

(1) The sample sizes are very small.

(2) The study only goes up to grade 3

(3) There is unusual variability in the grade 3 1989-90 score (standard deviation = 2.6). (The sd of 7.3 for comparisons in grade 1 appears to be an error; more likely, it was .73).

(4) The 1990-91 cohort shows a decline from grades 2 to 3, making less than one year progress in one year.

It is noteworthy that “Spanish amigos” children were behind comparisons on a measure of “communication skills” (done in an interview) in grade 1 but performed better than comparisons in grades 2 and 3, increasing their lead in grade 3.

Table 3
"Spanish Amigos" English Competence (grade level equivalents)

1989-1990 cohort				
Grade- n	score	comp n	comparison score	expected
1-8	1.38 (.17)	8	.95 (7.3)	1.7
2-10	2.07 (.71)	9	1.92 (.53)	2.7
3- 6	4.65 (2.6)	6	1.48 (.26)	3.7
1990-91 cohort				
1-12	1.29 (.43)	12	.69 (.49)	1.7
2-10	3.11 (1.16)	10	1.56 (.52)	2.7
3-7	2.87 (1.44)	7	2.14 (.47)	3.7

From: Cazabon, Lambert and Hall (1993)

In their Houston report, Thomas and Collier go a long way toward controlling for pre-existing differences among the groups. In their report, data is provided for two-way bilingual education as well as two alternative programs: transitional bilingual education (TBE) and “developmental bilingual education.” All three provide

equivalent amounts of Spanish and English up to the end of grade 3. In grade 4, the transitional program contains more English, while the other two maintain 50-50 Spanish-English.

Table 4
Comparison of Bilingual Programs/Cross-Sectional (NCE/Percentiles)

grade	TBE	Developmental	TBE	2-way
1	51.5/53	50.1/50	51.9/54	59.8/68
2	47.3/45	47.9/46	44.7/40	54.9/59
3	46.8/44	46.9/44	48.9/48	54/58
4	44.7/40	43.9/39	44.4/40	52.6/55
5	41.4/34	41.4/34	38.9/30	51.3/52

from: Thomas and Collier (2002)

Table 4 presents NCE scores and percentiles for English reading. Note that TBE groups appear twice in the table: Each type of "enriched" bilingual education program had its own comparison group, matching it in neighborhood and socioeconomic status.

Cross-sectional comparisons pit TBE comparison groups against each of the enriched bilingual options. There appears to be no difference in the TBE-developmental comparison, but two-way students do better than TBE comparisons. Also, two-way students clearly do better than the developmental students. Although this was not a “planned” comparison with control for socio-economic status, the result is noteworthy because two-way students performed better than TBE students, but developmental students did not.

Houston: Potential Confounds

Before concluding that this data demonstrates the superiority of two-way programs, we need to discuss some unusual aspects of the Houston data.

1. Scores for all groups are high in grade 1. In fact, they are higher than the level typically required for reclassification. This suggests that a significant percentage of the children had considerable English knowledge before starting school.
2. Scores then decline for all groups, as we saw before in Castillo (2001). By grade 5, in fact, scores for some groups are below typical reclassification levels. This suggests that higher scoring children are being exited and lower scoring children are entering the program late. (Note that if this were true, it would exaggerate the effectiveness of a program that kept all its students and did not allow new ones to enter late, e.g. two-way bilingual education.)
3. The two-way advantage was present early, in grade 1. This suggests that a selection bias was present, that the two-way students were exceptional from the beginning.

Thomas and Collier (personal communication) have pointed out that that children tested at the end of grade 1 could have been in school and receiving the benefits of bilingual education for three years before testing. Houston offers bilingual pre-kindergarten and kindergarten. It is thus possible that first grade scores do not represent beginning scores.

Thomas and Collier's Table C-8 provides some helpful information: It is a quasi-longitudinal analysis of children who have been in the Houston system for several years. It is not a pure longitudinal analysis, because different numbers of children were tested at each level; thus, the same cohorts were followed, but precisely the same children were not tested each year. Sometimes, in fact, the numbers vary quite a bit; there was a large increase in the number tested in grade 4 (see table 5). A valuable aspect of this data is the fact that separate scores are presented for children who began school in pre-k, kindergarten, and grade 1.

Table 5, from Thomas and Collier's Table C-8, presents scores of children who began in school in grade 1, who did not have pre-kindergarten or kindergarten. When tested at the end of grade 1, they had had only one year of schooling. In all cases I present the sample size in parentheses, followed by the NCE score given in Thomas and Collier (2002), followed by the corresponding percentile. Note that scores are quite high even when children start school at grade 1.

Table 5
Children Beginning School at Grade 1 (NCE/Percentiles)

grade	TBE	DEV	TBE	2-way
1	(25) 48.9/ 48	(15) 44.5/40	(9) 46.3/44	(11) 51.3/53
2	(24) 56.9/63	(18) 51.5/53	(8) 39.5/31	(8) 51.4/53
3	(34) 54.3/58	(20) 49.5/49	(14) 54/58	(15) 55.9/61
4	(169) 51.5/53	(169) 45.3/41	(40) 48.8/48	(18) 42/35

From: Thomas and Collier, 2002, table C-8.

With respect to concern (2), the decline, Thomas and Collier (personal communication) point out that declines are a common occurrence as students move toward middle school and a more demanding curriculum.

Thomas and Collier (personal communication) also point out that children are offered bilingual education in Texas from pre-kindergarten to grade five. Exit is not automatic when a certain level is reached. If this is so, if most students stay in the program, the exiting of high scorers is not a factor in the decline and two-way does not have an unfair advantage in this regard.

Interestingly, these drops appear to coincide with a reduction of the amount of instruction provided in the first language. They may also be related to a problem nearly all language minority children suffer from: because they also tend to be of low socio-economic class, they tend to live in neighborhoods and attend schools with little access to books (Neuman and Celano, 2001). Recreational reading may make a particularly strong contribution to reading tests at the upper grade levels.

With respect to concern (3), the early advantage of the two-way group, Thomas and Collier note that this group had the advantage of having two-way bilingual education in pre-kindergarten and kindergarten. Table 5, however, shows that their advantage was present even for those starting in grade 1.

As Thomas and Collier point out, the sample sizes in the longitudinal study are small, and one must be cautious in generalizing from them. Also, as noted earlier, the data is not strictly longitudinal. These concerns are, of course, an invitation for additional research.

Comparison with Developmental Bilingual Education

de Jong (2004) compared a two-way and a developmental bilingual program, focusing on children who stayed in the program continuously from kindergarten to

grade 5. The sample size was modest (26 in the two-way and 19 in the developmental program), but results were similar for children in other cohorts not (yet) followed to grade five. On the LAS reading test, two-way children did better than those in the developmental program, but the latter group appeared to be closing the gap. Scores are also quite high, but the test used was designed for English Learners, not native speakers of English. Because of the small sample and the presence of a ceiling effect (all scores very high), table 6 includes results of the LAS writing test as well. Again the two-way children do somewhat better.

The results probably underestimate the advantage of the two-way group: It appears that the developmental children knew more English when they came to school (table 6). On a measure of oral English comprehension, they scored very close to the proficient level after kindergarten (3.4 on a 5 point scale, when 4 and 5 = proficient in English).

Table 6
Two-Way Versus Developmental Bilingual Education

	READING		WRITING	
	2-way	develop.	2-way	develop.
2	82	68	69	52
3	94	82	88	72
4	92	88	83	72
5	97	95	87	82

	ORAL COMP.	
	2-way	develop.
k	2.2	3.4
1	3.8	3.8
2	4.3	4.6
3	4.2	4
4	4.3	4.8
5	5	4.8

from: de Jong (2004)

Group Four: Comparisons to Native Speakers of English

According to Alanis (2000), Spanish-dominant students in two schools in grades 3,4, and 5 in a two-way program did as well as English-dominant students in the same program and as well as monolingual English speakers on the Texas Learning Index test of English reading. The sample size was small: in one school, only seven Spanish-dominant children were tested, in the other, only 27. In addition, it was clear that the Spanish-dominant children did not begin school with zero English: Of 18 students tested in oral English, only two were considered "non- proficient" in grade 1. Thirteen were classified as "limited," two as "proficient" and one was at the "mastery" level in English.

The most serious problem with this study, however, was that it does not really qualify as a two-way bilingual program. Both English and Spanish-dominant children were in the same classroom, but the program had much more emphasis on English than Spanish. It appeared to be an English immersion program with some support in the primary language: it was supposed to be 50/50, but "it was clear from classroom observations and interviews that the 50/50 split was not implemented ... teachers ... utilized more Spanish than English and lacked Spanish resources in all content areas" (p. 16). Even though the schools were on the border (Brownsville), "most students had a strong preference for English" and "teachers stressed the English TAAS [Texas Assessment of Academic Skills] for their students" (p. 16). Alanis also notes that there was no emphasis placed on Spanish literacy: "the original goal of the program was only Spanish oral proficiency as opposed to equal levels of bilingualism and biliteracy" (p. 16).[2]

Some of the studies discussed previously also compared two-way students to English dominant and native speakers of English. Recall that Castillo (2001) found that two-ways did better than a group of native English speakers (n = 30), but only 11 two-way students were tested and there was no information on their initial level of English. Ajuria (1994)'s two-way students also did better than native speakers of English in the mainstream (n = 20), but recall that the mainstream classroom was severely lacking in many essentials and a modified version of a standardized test was used. Clayton (1993) reported that two-way students did better than a group of Native American children with similar backgrounds, but it is not clear whether the "two-way" children were in two-way or whether it was a combined sample of children in two-way and TBE: The sample size for the comparison with Native American children suggests it was a combined group: there were 83 children in the

two-way group and 59 in the TBE group in the two-way vs. TBE comparison, but 150 in the two-way vs. Native American comparison.

Thomas and Collier (2002) also provide a comparison with native speakers of English in their Houston sample. All native speakers performed around the 50th percentile in English reading; two-way students scored above this level at all grade levels tested, an impressive result even though there was no control for variables such as neighborhood or poverty.

Discussion

With respect to question (1), “Do they acquire significant amounts of English?, several studies reassure us and show that children in two-way programs improve in English. Table 7 presents grade five results in NCE’s and percentiles for the three studies using standardized tests, including all groups tested. Even discounting the lowest score as an outlier (9th percentile in Lindholm), the results are variable, and not always high enough for reclassification, and the Thomas and Collier score is much higher than the others.

Table 7
Attainment by Grade 5

study	scores: NRC (percentiles) at grade 5
de Jong (2002)	38 (29); 38 (29); 40 (30); 42 (35); 43 (37)
Lindholm (2002)	20 (9); 34 (22); 40 (30)
Collier & Thomas (2002)	51 (53)

With respect to question (2), whether two-way students do better than English learners in the mainstream, two studies suffer from very small samples, short durations, and lack of measures for initial competence. In another, Thomas and Collier’s Houston data, two-way students do better than those in submersion but not as well as those in ESL-only. This data, however, is uncontrolled for a number of important factors.

With respect to question (3), two-way children outperform children in transitional bilingual education in two studies, but do worse in another. None of the studies provides us assurance that the children were at similar levels when starting school,

one study, the Amigos project, ends at grade 2 and has a small sample size, and in the Houston study scores for all children were very high at grade 1, and are lower in subsequent years. In addition, children in two-way in Houston were superior and had high scores after only one year of school, suggesting a selection bias and considerable previous English competence. The only direct comparison of two-way with a developmental program showed evidence that two-way children did better, but sample sizes were modest and the developmental children appeared to be closing the gap between the two programs (note that Thomas and Collier, 2002. contains an indirect comparison of two-way with developmental bilingual education, with two-way doing better.)

Two-way children do as well or better than native speakers of English in five studies. In three of the five, the sample sizes were small or modest, one of the "two-way" programs was really immersion with some oral first language support, and durations were short in two studies.

Table 8
Summary of Studies

study	results	commentary
NO COMP. GROUP		
Sugarman & Howard	gains in English proficiency	tests unstandardized, no comparison group
		no measure of initial competence
de Jong (2002)	2-way achieve 29 to 37 percentile by grade 5	small n, no comparison group
		no measure of initial competence
Lindholm	2 -way achieve 23 to 30 percentile by grade 6	no comparsion group
	one group only at 9 percentile in grade 5	
Thomas/Collier (Oregon)	2-way better than state, district norms	small n (12), initial competence not clear

COMP. W. MAINSTREAM		
Ajuria	2-way better than comparisons in mainstream	small n (11,7) ,only 1 year, no gains
		second semester, comparisons in print-
		deprived environment, modified test
		no measure of initial competence
Castillo	2-way much better than comparisons in	small n (4 comparisons!), up to grade 2
	mainstream	only, 2-way scores very high after grade
		one, then decline, no measure of initial
		competence
Thomas/Collier (Houston)	ESL only > TBE > submersion	uncontrolled
COMP. W. TBE		
Clayton	TBE graduates do better than 2-way graduates	no measure of initial competence
		2-way had "inferior" reading program
Cazabon et. al.	2-way do better than TBE	small n, up to grade 2 only, initial
		English competence unclear
Thomas/Collier (Houston)	2-way better than TBE	grade 1 scores very high, then decline

		no measure of initial competence
		2-way advantage present very early
COMP W. DEVELOPMENTAL		
de Jung (2004)	2-way better than developmental	small n, ceiling on reading test
		Dev. have better oral compr. at K
COMP. W. Nspeakers of Eng		
Alanis	2-way = English dominant	small n, 2-way did not begin with zero
		English; 2-way not really 2-way
Ajuria	2-way better than English native speakers	test modified, mainstream class print-deprived
		one year only
Castillo	2-way better than English native speakers	up to grade 2 only; no measure of intial
		competence; small n (11 2-way children)
Clayton	2-way better than Native American children	not clear if 2-way group alone was
		included or entire 2-way + TBE group
Thomas/Collier (Houston)	2 way better than English native speakers	Uncontrolled (see also above)

The results are thus encouraging but they are not the overwhelming and massive

support we sometimes read about in the popular press.

It should be pointed out that supporters of bilingual education have criticized studies claiming to support immersion for similar flaws. In Krashen (1996) I faulted Gersten (1985) for a small sample size and a short duration (up to the end of grade 2), I also noted that Rossell's (1990) analysis was suspect because scores were very high long before reclassification.

Rossell and Baker (1966) criticize many studies for not controlling for individual differences at the outset of the program, which is also a problem in the studies discussed here. I argued (Krashen, 1996) that with large numbers of studies, this is not a problem if there is no reason to suspect differences; large numbers of studies provide quasi-randomization. This is clearly what is called for in the case of two-way bilingual education.

This review has obvious limitations.

I have limited this analysis to performance of English learners on tests of academic English. While we need to be reassured that English language development is satisfactory, we also need to consider long term cognitive development, social and attitudinal factors, the ease of implementation and efficacy of different versions of two-way bilingual education, the effect on the heritage language and the effect on majority language students, especially those from low-income families who may have little opportunity for first language development outside of school because of print-poor environments.

One could argue that one need not show superiority in English language development: If children are clearly showing substantial growth in English, enough to access the regular curriculum in a reasonable amount of time, small differences among programs are clearly unimportant. If all students eventually acquire English very well, if in fact it turns out that students in one program reach a certain level of English a few months faster than children in another clearly does not matter. But thus far, the attainment of children in two-way programs in academic English is not consistently overwhelming.

In addition, I have ignored the variants among two-way programs in design. Lindolm (2001), for example, presents two different models. With more studies, we will be able to consider design as a predictor.

Conclusions

Only a handful of studies exist, and they report generally positive but variable attainment in academic English among English learners. In studies comparing two-way children with those in other options, sample sizes are often small, there is usually no control for initial differences, and scores are sometimes high at the beginning and then decline.

Supporters of bilingual education, such as this writer, have critiqued studies claiming to support immersion on similar grounds.

Thus, a close look at the data shows that two-way programs show some promising results, but research has not yet demonstrated that they are the best possible program.

Acknowledgments: I thank Wayne Thomas, Virginia Collier, James Crawford, and Giselle Waters for comments on earlier versions of this paper.

References

Ajuria, A. (1994). An exploration of classroom activity and student success in a two-way bilingual and a mainstream program. Ph Dissertation, Boston College.

Alanis, I. (2000). A Texas two-way bilingual program: Its effects on linguistic and academic achievement. *Bilingual Research Journal,* 24 (3).

Castillo, C. (2001). The effects of dual language education program on student achievement and development of leadership abilities. Ph. D dissertation, Our Lady of the Lake University.

Cazabon, M., Lambert, W., & Hall, G. (1993). Two-way bilingual education: A progress report on the Amigos program. Santa Cruz, CA and Washington, DC: National Center for Research on Cultural Diversity and Second Language Learning.

Clayton, S. 1993. Language policy and equal status: A change from monolingualism to bilingualism in a small district in California. Ph D dissertation, Claremont Graduate School and San Diego State University

Christian, D., Montone, C., Lindholm, K. & Carranza, I. (1997). Profiles in two-way immersion education. McHenry, IL: Delta Systems.

Coy, S. & Litherland, L. (2000). From a foreign language perspective: A snapshot view of a dual language program in two inner-city high poverty elementary schools. ERIC Document ED 446 450.

de Jong, E. (2002). Effective bilingual education: From theory to academic achievement in a two-way biingual program. *Bilingual Research Journal*, 26 (1).

de Jong, E. (2004). L2 proficiency development in a two-way and a developmental bilingual program. *NABE Journal of Research and Practice*, 2(1).

Gersten, R. (1985). Structured immersion for language minority students: Results of a longitudinal evaluation. *Educational Evaluation and Policy Analysis* 7. 187- 196.

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

Lindholm, K. & Fairchild, H. (1990). Evaluation of an elementary school bilingual immersion program. In Padilla, A., Fairchild, H., & Valadez, C. (Eds.) *Bilingual education: Issues and strategies.* (pp. 126-136). Beverly Hills: Sage.

Lindholm, K. (2002). *Dual language education.* Clevedon, UK: Multilingual Matters

Neuman, S., & Celano. D. (2001). Access to print in low-income and middle-income communities. *Reading Research Quarterly*, 36 (1), 8-26.

Oller, D.K. & Eilers, R. (2002) Language and literacy in bilingual children. Clevedon, UK: Multilingual Matters.

Perez. B. & Flores, B. (2002). Biliteracy development in two-way immersion classrooms: Analysis of third grade Spanish and English reading. In Schallert, D., Fairbanks, C., Worthy, J., Maloch, B. & Hoffman, J. (Eds). *51st yearbook of the National Reading Conference* (pp. 357-367). Oak Creek, WI: National Reading Conference.

Rossell, C. (1990). The effectiveness of educational alternatives for limited-English-proficient children. In G. Imhoff (Ed.) *Learning in Two Languages* (pp. 71-121). New Brunswick, NJ: Transaction Publishers.

Rossell, C. & Baker, K., (1996). The educational effectiveness of bilingual education. *Research in the Teaching of English*, 30, 7-74.

Senesac, B. (2002). Two-way bilingual immersion: A portrait of quality schooling. *Bilingual Research Journal* 26 (1), 1-17.

Sugarman, J. & Howard, L. 2001. Two-way immersion shows promising results: Findings from a new study. Washington, DC: Center for Applied Linguistics.

Thomas, W. P., & Collier, V. (1997). School effectiveness for language minority students. Washington, DC: NCBE. Available: http://www.ncela.gwu.edu/ncbepubs/resource/effectiveness/thomas-collier97.pdf.

Thomas, W. P. & Collier, V. (2002). A national study of school effectiveness for language minority students' long-term academic achievement. Final Report: Project 1.1. University of California, Santa Cruz: Center for Research, Education, Diversity and Excellence. http://www.crede.ucsc.edu/research/llaa/1.1_final.html

Endnotes

1. In Christian, Montone, Lindholm, and Caarraza (1997), test scores were combined for minority and majority language children. In addition, many of the Spanish speakers entered school with significant English competence; 45% of entering students at one school (Inter-American) were considered “bilingual” and half of the Spanish speakers at another (River Glen) were classified as fluent in English in grade one. Lindholm and Fairchild (1990) cover progress only to grade one and the comparison group included English and Spanish dominant children. Senesac (2002) reported that Spanish speakers in a two-way program performed at grade level in grades three through eight, but approximately half the sample were never classified as limited English proficient (Senesac, 2002, p. 4).

2. Stipek, Ryan and Alarcon (2001) is another “two-way” program that had tremendous emphasis on English: For children in preK and K, teacher talk was in English 57% of the time to Spanish-dominant children, mixed 19%, and in Spanish 24% of the time. For grades 1 and 2, teacher input was in English 77% of the time, 9% mixed, and 14% in Spanish. Pre-K and K Spanish-speaking children responded in Spanish (with no mixing) only 11% of time, and those in grades 1 and 2, only 12%