

Running head: LANGUAGE OF INSTRUCTION

A Synthesis of Research on Language of Reading Instruction for English Language Learners

Robert E. Slavin
Johns Hopkins University
200 W. Towsontown Blvd.
Baltimore, MD 21204
rslavin@successforall.net
410 616 2310

Alan Cheung
Success for All Foundation
200 W. Towsontown Blvd.
Baltimore, MD 21204
acheung@successforall.net
410 616 2318

Revision
July, 2004

This paper was written under funding from the Institute of Education Sciences, U.S. Department of Education (Grant No. OERI-R-117-40005). However, any opinions expressed do not necessarily reflect IES positions or policies.

We would like to thank Margarita Calderón, Diane August, Tim Shanahan, Isabel Beck, Jay Greene, Nancy Madden, and Bette Chambers for their comments on an earlier draft, and Susan Davis and the Center for Applied Linguistics for assistance with the literature search.

Running head: LANGUAGE OF INSTRUCTION

**A Synthesis of Research
on Language of Reading Instruction
for English Language Learners**

**Revision
July, 2004**

Abstract

This article reviews experimental studies of reading programs for English language learners, focusing on comparisons of various bilingual and English-only programs. The review method is best-evidence synthesis, which uses a systematic literature search, quantification of outcomes as effect sizes, and extensive discussion of individual studies that meet inclusion standards. A total of 18 studies met the inclusion standards. Among 13 studies focusing on elementary reading for Spanish-dominant students, 9 favored bilingual approaches on English reading measures, and 4 found no differences, for a median effect size of +0.52 (based on 8 studies with sufficient data for computation of ES). Two of three studies of heritage languages (French, Choctaw, and Cherokee) and two secondary studies favored bilingual approaches. The review concludes that while the number of high-quality studies is small, existing evidence favors bilingual approaches, especially paired bilingual strategies that teach reading in the native language and English at different times each day. Research using longitudinal, randomized designs is needed to understand how best to ensure reading success for all English language learners.

The reading education of English language learners (ELLs) has become one of the most important issues in all of educational policy and practice. As the pace of immigration to the U.S. and other developed countries has accelerated in recent decades, increasing numbers of children in U.S. schools come from homes in which English is not the primary language spoken. As of 1999, 14 million Americans ages 5-24, or 17% of this age group, spoke a language other than English at home. This is more than twice the number of such individuals in 1979, when only 9% of Americans ages 5-24 spoke a language other than English at home (NCES, 2004). While many children of immigrant families succeed in reading, too many do not. In particular, Latino and Caribbean children are disproportionately likely to perform poorly in reading and in school. As No Child Left Behind and other federal and state policies begin to demand success for all subgroups of children, the reading achievement of English language learners is taking on even more importance. Thousands of schools cannot meet their adequate yearly progress goals, for example, unless their English language learners are doing well in reading. More importantly, American society cannot achieve equal opportunity for all if its schools do not succeed with the children of immigrants.

Sixty-five percent of non-English speaking immigrants in the U.S. are of Hispanic origin (NCES, 2004), and this is also one of the fastest growing of all groups. Hispanics have recently surpassed African Americans as the largest minority group in the U.S. Hispanic students as a whole, including English proficient children in the second generation and beyond, score significantly lower in reading than other students. On the National Assessment of Educational Progress (NAEP; Grigg, Daane, Jin, & Campbell, 2003), which excludes children with the lowest levels of English proficiency from

testing, only 44% of Latino fourth graders scored at or above the “basic” level, in comparison to 75% of Anglo students. Only 15% of Latino fourth graders scored at “proficient” or better compared to 41% of Anglos. Further, 31% of students who speak Spanish at home fail to complete high school, compared to 10% of students who speak only English (NCES, 2004).

There is considerable controversy, among policymakers, researchers, and educators, about how best to ensure the reading success of English language learners. While there are many aspects of instruction that are important in the reading success of English language learners, one question has dominated all others: What is the appropriate role of the native language in the instruction of English language learners? In the 1970s and 1980s, policies and practice favored bilingual education, in which children were taught partially or entirely in their native language, and then transitioned at some point during the elementary grades to English-only instruction. Such programs are still widespread, but from the 1990s to the present, the political tide has turned against bilingual education, and California, Arizona, Massachusetts, and other states have enacted policies to greatly curtail bilingual education. Recent federal policies are restricting the amount of time children can be taught in their native language. Among researchers, the debate between advocates of bilingual and English-only reading instruction has been fierce, and ideology has often trumped evidence on both sides of the debate (Hakuta, Butler, & Witt, 2000).

This article reviews research on the language of effective reading instruction for English language learners in an attempt to apply consistent, well-justified standards of evidence to draw conclusions about the role of native language in reading instruction for

these children. The review applies a technique called “best-evidence synthesis” (Slavin, 1986), which attempts to use consistent, clear standards to identify unbiased, meaningful information from experimental studies and then discusses each qualifying study, computing effect sizes but also describing the context, design, and findings of each study. Best-evidence synthesis closely resembles meta-analysis, but it requires more extensive description of key studies. Details of this procedure are described below. The purpose of this review is to examine the evidence on language of instruction in reading programs for English language learners to discover how much of a scientific basis there is for competing claims about effects of various bilingual programs, to inform practitioners, policymakers, and researchers about the current state of the evidence on this topic as well as gaps in the knowledge base in need of further scientific investigation.

Language of Instruction

For many years, the discussion about effective reading programs for English language learners has revolved around the question of the appropriate language of instruction for children who speak languages other than English. Proponents of native language instruction argue that while children are learning to speak English, they should be taught to read in their native language first, to avoid the failure experience that is likely if children are asked to learn both oral English and English reading at the same time. Children are then transitioned to English-only instruction when their English is sufficient to ensure success, usually in third or fourth grade. Alternatively, many programs teach young children to read both in their native language and in English at different times of the day or on alternating days. There is a great deal of evidence that

children's reading proficiency in their native language is a strong predictor of their ultimate English reading performance (Garcia, 2000; Lee & Schallert, 1997; Reese, Garnier, Gallimore, & Goldenberg, 2000), and that bilingualism itself does not interfere with performance in either language (Yeung, Marsh, & Suliman, 2000). Advocates also argue that without native language instruction, English language learners are likely to lose their native language proficiency, or fail to learn to read in their native language, losing skills that are of economic and social value in the world today. Opponents, on the other hand, argue that native language instruction interferes with or delays English language development, and relegates children who receive such instruction to a second-class, separate status within the school and, ultimately, within society. They reason that more time on English reading should translate into more learning (see Rossell & Baker, 1996).

Reviews of the educational outcomes of native language instruction have reached sharply conflicting conclusions. In a meta-analysis, Willig (1985) concluded that bilingual education was more effective than English-only instruction. Wong-Fillmore & Valadez (1986) came to the same conclusion. However, Rossell & Baker (1996) came to the opposite conclusion, claiming that most methodologically adequate studies found bilingual education to be no more effective than English-only programs. Greene (1997) re-analyzed the studies cited by Rossell & Baker and reported that many of the studies they cited lacked control groups, mischaracterized the treatments, or had other serious methodological flaws. Among the studies that met an acceptable standard of methodological adequacy, including all of the studies using random assignment to conditions, Greene found that the evidence favored programs that made significant use of

native language instruction. August & Hakuta (1997) concluded that while research generally favored bilingual approaches, the nature of the methods used and the populations to which they were applied were more important than the language of instruction per se. Quantitative research on the outcomes of bilingual education has diminished in recent years, but policy and practice are still being influenced by conflicting interpretations of research on this topic. The following sections systematically examine this evidence to attempt to discover what we can learn from research to guide policies in this controversial arena.

English Immersion and Bilingual Programs

When a child enters kindergarten or first grade with limited proficiency in English, the school faces a serious dilemma. How can the child be expected to learn the skills and content taught in the early grades while he or she is learning English? There may be many solutions, but two fundamental categories of solutions have predominated: *English immersion* and *bilingual education*.

English Immersion. In immersion strategies, English language learners are expected to learn in English from the beginning, and their native language plays little or no role in daily reading lessons. Formal or informal support is likely to be given to ELLs to help them cope in an all-English classroom. This might or might not include help from a bilingual aide who provides occasional translation or explanation, a separate English as a Second Language class to help build oral English skills, or use of a careful progression from simplified English to full English as children's skills grow. Teachers of English language learners might use language development strategies, such as total physical

response (acting out words) and realia (concrete objects to represent words), to help them internalize new vocabulary. They might simplify their language and teach specific vocabulary likely to be unfamiliar to ELLs (see Calderón, 2001; Carlo et al., in press). Immersion may involve placing English language learners immediately in classes containing English monolingual children, or it may involve a separate class of ELLs for some time until children are ready to be mainstreamed. These variations may well have importance in the outcomes of immersion strategies, but their key common feature is the exclusive use of English texts, with instruction overwhelmingly or entirely in English.

Many authors have made distinctions among different forms of immersion. One term often encountered is “submersion,” primarily used pejoratively to refer to “sink or swim” strategies in which no special provision is made for the needs of English language learners. This is contrasted with “structured English immersion,” which refers to a well-planned, gradual phase-in of English instruction relying initially on simplification and vocabulary-building strategies. In practice, immersion strategies are rarely pure types, and in studies of bilingual education, immersion strategies are rarely described beyond their designation as the English-only “control group.”

Bilingual Education. Bilingual education differs fundamentally from English immersion in that it gives English language learners significant amounts of instruction in reading and/or other subjects in their native language. In the U.S., the overwhelming majority of bilingual programs involve Spanish, due to the greater likelihood of a critical mass of students who are Spanish-dominant and to the greater availability of Spanish materials than those for other languages. There are bilingual programs in Portuguese, Chinese, and other languages, but these are rare. In transitional bilingual programs,

children are taught to read entirely in their native language through the primary grades and then transition to English reading instruction somewhere between second and fourth grade. English oracy is taught from the beginning, and subjects other than reading may be taught in English, but the hallmark of transitional bilingual education is the teaching of reading in the native language for a period of time. Such programs can be “early-exit” models, with transition to English completed in second or third grade, or “late-exit” models, in which children may remain throughout elementary school in native-language instruction to ensure their mastery of reading and content before transition (see Ramirez, Pasta, Yuen, Billings, & Ramey, 1991). Alternatively, “paired bilingual” models teach children to read in both English and their native language at different time periods each day or on alternating days. Within a few years, the native language reading instruction may be discontinued, as children develop the skills to succeed in English. Willig (1985) called this model “alternative immersion,” because children are alternatively immersed in native language and English instruction.

Two-way bilingual programs, also called dual language or dual immersion, provide reading instruction in the native language (usually Spanish) and in English both to ELLs and to English speakers (Calderón & Minaya-Rowe, 2003; Howard, Sugarman, & Christian, 2003). For the ELLs, a two-way program is like a paired bilingual model, in that they learn to read both in English and in their native language at different times each day.

A special case of bilingual education is programs designed more to preserve or show respect for a given language than to help children who are genuinely struggling with English. For example, Bacon, Kidd, & Seaberg (1982) studied a Cherokee language

program used with children who were of Cherokee ancestry but in many cases did not speak the language. Morgan (1971) studied a program in Louisiana for children whose parents often spoke French at home, but generally spoke English themselves. Such “heritage language” programs are included in this review if the outcome variable in the study is an English reading measure. They should be thought of, however, as addressing a different problem from that addressed by bilingual or immersion reading instruction for children who are limited in English proficiency.

Problems of Research on Language of Instruction.

Research on the achievement effects of teaching in the child’s native language in comparison to teaching in English suffers from a number of inherent problems beyond those typical of other research on educational programs. First, there are problems concerning the ages of the children involved, the length of time they have been taught in their first language, and the length of time they have been taught in English. For example, imagine that a transitional bilingual program teaches Spanish-dominant students primarily in Spanish in grades K-2, and then gradually transitions them to English by fourth grade. If this program is compared to an English immersion program, at what grade level is it legitimate to assess the children in English? Clearly, a test in second grade is meaningless, as the bilingual children have not yet been taught to read in English. At the end of third grade, the bilingual students have partially transitioned, but have they had enough time to become fully proficient? Some would argue that even the end of fourth grade would be too soon to assess the children fairly in such a comparison,

as the bilingual children need a reasonable time period in which to transfer their Spanish reading skills to English (see, for example, Hakuta, Butler, & Witt, 2000).

A related problem has to do with pretesting. Imagine that a study of a K-4 transitional Spanish bilingual program began in third grade. What pretest would be meaningful? An English pretest would understate the skills of the transitional bilingual students, while a Spanish test would understate the skills of the English immersion students. For example, Valladolid (1991) compared gains from grades 3 to 5 for children who had been in either bilingual or immersion programs since kindergarten. These children's "pretest" scores are in fact posttests of very different treatments. Yet studies comparing transitional bilingual and immersion programs are typically too brief to have given the students in the transitional bilingual programs enough time to have fully transitioned to English. In addition, many studies begin after students have already been in bilingual or immersion treatments for several years.

The studies that do look at four- or five-year participations in bilingual or immersion programs are usually retrospective (i.e., researchers search records for children who have already been through the program). Retrospective studies also have characteristic biases, in that they begin with the children who ended up in one program or another. For example, children who are removed from a given treatment for systematic reasons, such as Spanish-dominant students removed from English immersion because of their low performance there, can greatly bias a retrospective study, making the immersion program look more effective than it was in reality.

Many inherent problems relate to selection bias. Children end up in transitional bilingual education or English immersion by many processes that could be highly

consequential for the outcomes. For example, Spanish-dominant students may be assigned to Spanish or English instruction based on parent preferences. Yet parents who would select English programs are surely different from those who would select Spanish in ways that would matter for outcomes. A parent who selects English may be more or less committed to education, may be less likely to be planning to return to a Spanish-speaking country, or may feel very differently about assimilation. Thomas & Collier (2002) reported extremely low scores for Houston students whose parents refused to have their children placed in either bilingual or English as a Second Language programs. Are those scores due to relatively positive effects of bilingual and ESL programs, or are there systematic differences between children whose parents refused bilingual or ESL programs and other children? It is impossible to say, as no pretest scores were reported.

Bilingual programs are more likely to exist in schools with very high proportions of English language learners, and this is another potential source of bias. For example, Ramirez et al. (1991) found that schools using late-exit bilingual programs had much higher proportions of ELLs than did early-exit bilingual schools, and English immersion schools had the smallest proportion of ELLs. This means that whatever the language of instruction, children in schools with very high proportions of ELLs are conversing less with native English speakers both in and out of school than might be the case in an integrated school and neighborhood that uses English for all students because its proportion of ELLs is low. Most problematically, individual children may be assigned to native language or English programs because of their perceived or assessed competence. Native language instruction is often seen as an easier, more appropriate placement for ELLs who are struggling to read *in their first language*, while students who are very

successful readers in their first language or are felt to have greater potential are put in English-only classes. This selection problem is most vexing at the point of transition, as the most successful students in bilingual programs are transitioned earlier than the least successful children. A study of bilingual vs. immersion programs involving third or fourth graders may be seriously biased by the fact that the highest-achieving bilingual students may have already been transitioned, so the remaining students are the lowest achievers.

A source of bias not unique to studies of bilingual education but very important in this literature is the “file drawer” problem, the fact that studies showing no differences are less likely to be published or to otherwise come to light. This is a particular problem in studies with small sample sizes, which are very unlikely to be published if they show no differences. The best antidote to the “file drawer” problem is to search for dissertations and technical reports, which are more likely to present their data regardless of their findings (see Cooper, 1998).

Finally, studies of bilingual education often say too little about the bilingual and immersion programs themselves or the degree or quality of implementation of these programs. Yet bilingual models can vary substantially in quality, amount of exposure to English in and out of school, teachers’ language facility, time during the school day, instructional strategies unrelated to language of instruction, and so on.

Because of these inherent methodological problems, an adequate study comparing bilingual and immersion approaches would: a) randomly assign a large number of children to be taught in English or their native language; b) pretest them in their native language when they begin to be taught differentially, either in their native language or in

English (typically kindergarten); c) follow them long enough for the latest-transitioning children in the bilingual condition to have completed their transition to English and have been taught long enough in English to make a fair comparison; and d) collect data throughout the experiment to document the treatments received in all conditions.

Unfortunately, only a few, very small studies of this kind have ever been carried out. As a result, the studies that compare bilingual and English-only approaches must be interpreted with great caution.

Review Methods

This section focuses on research comparing immersion and bilingual reading programs applied with English language learners, with measures of English reading as the outcomes. The review uses a quantitative synthesis method called “best-evidence synthesis” (Slavin, 1986). It uses the systematic inclusion criteria and effect size computations typical of meta-analyses (see Cooper, 1998; Cooper & Hedges, 1994), but discusses the findings of critical studies in a form more typical of narrative reviews. This strategy is particularly well-suited to the literature on reading programs for English language learners, because the studies are few in number and are substantively and methodologically diverse. In such a literature, it is particularly important to learn as much as possible from each study, not just to average quantitative outcomes and study characteristics.

Literature Search Strategy

The literature search benefited from the assistance of the federally commissioned National Literacy Panel on the Development of Literacy Among Language Minority Children and Youth, chaired by Diane August and Timothy Shanahan. The first author was initially a member of the Panel, but resigned to avoid a two-year delay in publication of the present article. This review, however, is independent of the panel's report, and uses different review methods and selection criteria. Research assistants searched ERIC and other databases for all studies involving language minority students, English language learners, and related descriptors. Citations in other reviews and articles were also obtained. From this set, we selected studies that met the criteria described below.

Criteria for Inclusion

The best-evidence synthesis focused on studies that met minimal standards of methodological adequacy and relevance to the purposes of the review. These were as follows.

1. The studies compared children taught reading in bilingual classes to those taught in English immersion classes, as defined above. Studies of alternative reading programs for English language learners that held constant the language of instruction are discussed in a later section of this review.
2. Either random assignment to conditions was used, or pretesting or other matching criteria established the degree of comparability of bilingual and immersion groups before the treatments began. If these matching variables were not identical at pretest, analyses adjusted for pretest differences or data permitting such

adjustments were presented. Studies without control groups, such as pre-post comparisons or comparisons to “expected” scores or gains, were excluded. Studies with pretest differences exceeding one standard deviation were excluded. Those with pretest differences less than one standard deviation were included if they carried out appropriate statistical adjustments.

A special category of studies were rejected based on the requirement of pretest measurement before treatments began. These are studies in which the bilingual and immersion programs were already under way before pretesting or matching. For example, Danoff, Coles, McLaughlin, & Reynolds (1978), in a widely cited study, compared one-year reading gains in many schools using bilingual or immersion methods. The treatments began in kindergarten or first grade, but the pretests (and later, posttests) were administered to children in grades 2-6. Because the bilingual children were primarily taught in their native language in K-1, their pretests in second grade would surely have been affected by their treatment condition. Similarly, several studies tested children in upper elementary or secondary grades who had experienced bilingual or immersion programs in earlier years. These were included if premeasures were available from before the programs began, but in most cases such premeasures are not reported (see, for example, Thomas & Collier, 2002; Curiel, Stenning, & Cooper-Stenning, 1980).

3. The subjects were English language learners in elementary or secondary schools in English-speaking countries. Studies that mixed ELLs and English monolingual students in a way that does not allow for separate analyses were excluded (e.g., Skoczytas, 1972). Studies of children learning a foreign language were not

- included. However, Canadian studies of French immersion have been widely discussed, and are therefore discussed in a separate section.
4. The dependent variables included quantitative measures of English reading performance, such as standardized tests and informal reading inventories. If treatment-specific measures were used, they were included only if there was evidence that all groups focused equally on the same outcomes. Measures of outcomes related to reading, such as language arts, writing, and spelling, were not included.
 5. The treatment duration was at least one school year. For the reasons discussed earlier, even one-year studies of transitional bilingual education are insufficient, because students taught in their native language are unlikely to have transitioned to English. Studies even shorter than this do not address the question in a meaningful way.

Limitations

It is important to note that the review methods applied in this best-evidence synthesis have some important limitations. First, in requiring measurable outcomes and control groups, the synthesis excludes case studies and qualitative studies. Many such descriptions exist, and these are valuable in suggesting programs or practices that might be effective. Description alone, however, does not indicate how much children learned in a given program, or what they would have learned had they not experienced that program. Second, it is possible that a program that has no effect on reading achievement measures might nevertheless increase children's interest in reading or reading behaviors

outside of school. However, studies rarely measure such outcomes in any systematic or comparative way, so we can only speculate about them. Finally, it is important to note that many of the studies reviewed took place many years ago, and that both social and political contexts, as well as bilingual and immersion programs, have changed, so it cannot be taken for granted that outcomes described here would apply to outcomes of bilingual and immersion programs today.

Computation of Effect Sizes

If possible, effect sizes were computed for each study. An effect size is the experimental mean minus the control mean divided by the control group's standard deviation. When information was lacking, however, effect sizes were estimated using pooled standard deviations, exact *t*'s or *p* values, or other well-established estimation methods (see Cooper, 1998; Cooper & Hedges, 1994). If effect sizes could not be computed in a study that otherwise qualified for inclusion, the findings were still reported. No study was excluded solely on the grounds that it did not provide sufficient information for computation of an effect size.

Previous Quantitative Reviews

The debate about empirical research on language of instruction for English language learners has largely pitted two researchers, Christine Rossell and Keith Baker, against several other reviewers. Rossell & Baker have carried out a series of reviews and critiques arguing that research does not support bilingual education (see Baker & de Kanter, 1981, 1983; Baker, 1987; Rossell, 1990; Rossell & Ross, 1986). The most

comprehensive and recent version of their review was published in 1996. In contrast, Willig (1985) carried out a meta-analysis and concluded that research favored bilingual education, after controls were introduced for various study characteristics. Other reviewers using narrative methods have agreed with Willig, e.g., Wong-Fillmore & Valadez (1986). Baker (1987) and Rossell & Baker (1996) criticized the Willig (1985) review in detail and Willig (1987) responded to the Baker (1987) criticisms.

In a review commissioned by the Tomas Rivera Center, Jay Greene (1997) carefully re-examined the Rossell & Baker (1996) review. While Rossell & Baker used a “vote-counting” method in which they simply counted the numbers of studies that favored bilingual, immersion, or other strategies, Greene (1997) carried out a meta-analysis in which each study produced one or more effect sizes, the proportion of a standard deviation separating bilingual and English programs. Greene categorized only 11 of the 72 studies cited by Rossell & Baker as methodologically adequate, but among these he calculated an effect size of +0.21 favoring bilingual over English-only approaches on English reading measures. Among five studies using random assignment, Greene calculated an effect size of +0.41 on English reading measures.

As part of this review, we attempted to obtain every study reviewed by Rossell & Baker and by Willig, and independently reviewed each one against the consistent set of standards outlined previously. Consistent with Greene, we found that the Rossell & Baker (1996) review accepted many studies that lacked adequate methodology. Appendix 1 lists all of the reading studies cited by Rossell & Baker according to categories of methodological adequacy outlined in this article, which closely follow Greene’s categorization. As is apparent from the Appendix, only a few of the studies met

the most minimal of methodological standards, and most violated the inclusion criteria established by Rossell & Baker (1996) themselves. We found, however, that most of the 16 studies cited by Willig also do not meet these minimal standards. These are also noted in Appendix 1. In itself, this does not mean that the overall conclusions of either review are incorrect, but it does mean that the question of effects of language of instruction on reading achievement must be explored with a different set of studies than the ones cited by either Rossell & Baker or Willig. The Rossell & Baker and Willig studies can be categorized as follows (following Greene, 1997):

1. Methodologically adequate studies of elementary reading. These are studies that compared English language learners taught to read using bilingual or immersion strategies, with random assignment or well-documented matching on pretests or other important variables.
2. Methodologically adequate studies of secondary programs. We put two secondary school studies (Covey, 1973; Kaufman, 1968) in a separate category.
3. Canadian studies of French immersion. Several studies (e.g., Lambert & Tucker, 1972; Genesee & Lambert, 1983) evaluated French immersion programs in Canada. However, since they compared immersion to monolingual English instruction or to brief French-as-a-second-language classes, these are not evaluations of bilingual education.
4. Studies in which the target language was not the societal language. In addition to Canadian studies of French immersion in non-francophone areas (e.g., Day & Shapson, 1988), Ramos, Aguilar, & Sibayan (1967) studied various strategies for teaching English in the Phillipines.

5. Studies of outcomes other than reading. A few studies (e.g., Lum, 1971; Legarreta, 1979) assessed only oral language proficiency, not reading.
6. Studies in which pretesting took place after treatments were under way. As noted earlier, many studies (e.g., Danoff et al., 1978; Rosier & Holm, 1980; Rossell, 1990; Thomas & Collier, 2002; Valladolid, 1991) compared gains made in bilingual and immersion programs after the programs were well under way. Both Willig and Rossell & Baker included such studies, and Greene (1997) accepted them as “methodologically adequate,” but we would argue that they add little to understanding the effects of bilingual education.
7. Redundant studies. Rossell & Baker included many studies that were redundant with other studies in their review. For example, one longitudinal study (El Paso, 1987, 1990, 1992) issued three reports on the same experiment, but it was counted as three separate studies. Curiel’s 1979 dissertation was published in 1980, yet both reports are counted.
8. No evidence of initial equality. Several studies either lacked data on initial achievement, before treatments began, or presented data indicating pretest differences in excess of one standard deviation.
9. No appropriate comparison group. Many of the studies included by Rossell & Baker had no control group. For example, Burkheimer, Conger, Dunteman, Elliott, & Mowbray (1989) and Gersten (1985) used statistical methods to estimate where children should have been performing and then compared this estimate to their actual performance. Rossell & Baker’s own standards required “a comparison group of LEP students of the same ethnicity and similar language

background,” yet they included many studies that did not have such comparison groups. Further, many studies included by Rossell & Baker lacked any information about the initial comparability of children who experienced bilingual or English-only instruction (e.g., Matthews, 1979). This includes studies that retroactively compared secondary students who had participated in bilingual or English-only programs in elementary schools but failed to obtain measures of early academic ability or performance (e.g., Powers, 1978; Curiel et al., 1980). Other studies compared obviously non-comparable groups. As an example of the latter, Rossell (1990) compared one-year gains of English language learners in Berkeley, California, who were in Spanish bilingual or immersion programs, yet 48% of the ELLs, all in the immersion programs, were Asian, while all students in the Spanish bilingual program (32% of the sample) were, of course, Latino. Also, Legarreta (1979) compared Spanish-dominant children in bilingual instruction to mainly English-dominant children taught in English. Finally, Carlisle & Beeman (2000) compared Spanish-dominant children taught 80% in Spanish and 20% in English to those taught 80% in English and 20% in Spanish, so there was no English-only comparison group.

10. Brief studies. A few studies cited by Rossell & Baker involved treatment durations less than one year. For the reasons discussed earlier, studies of bilingual education lasting only 10 weeks (Layden, 1972) or four months (Balasubramonian, Seelye, & de Weffer, 1973) are clearly not relevant. Also, all but one of these brief studies also failed to meet inclusion standards on other criteria as well (e.g., they lacked pretests or had outcomes other than reading).

The Present Review

This review carries out a best-evidence synthesis of studies comparing bilingual and English approaches to reading in the elementary and secondary grades that meet the inclusion criteria outlined above. These include the methodologically adequate studies cited in the Willig (1985), Rossell & Baker (1996), and Greene (1997) reviews, as well as other studies located in an exhaustive search of the literature, as described previously. The characteristics and findings of these studies are summarized in Table 1.

=====
TABLE 1 HERE
=====

Studies of Beginning Reading for Spanish-Dominant Students

The largest number of studies focused on teaching reading to Spanish-dominant students in the early elementary grades. Thirteen studies of this kind met the inclusion criteria.

Three categories of bilingual programs were distinguished. The most common among the qualifying studies were studies of paired bilingual strategies, in which students were taught to read in English and in Spanish at different times of the day, beginning in kindergarten or first grade and continuing through the end of the study. Pairing may not begin on the first day of the school year, but if children are being taught both Spanish and

English reading during their first year of reading instruction, the program is considered a paired model. A second category involved evaluations of programs in which children were taught reading in Spanish for one year before a transition to paired bilingual instruction (English and Spanish). A third category consisted of a single study, the well-known Ramirez, Yuen, & Ramey (1991) longitudinal evaluation of transitional bilingual education. In that study, the experimental group was taught in Spanish in kindergarten and first grade and then transitioned to English during second grade, completing the transition by the end of second grade. Ironically, this treatment was referred to as “early transition” by Ramirez et al., but among the studies meeting the inclusion criteria, this was the latest transition located. A “late transition” treatment was also studied by Ramirez et al., but this comparison did not meet the inclusion criteria due to lack of an appropriate control group (see the study description below). Finally, one study, by Saldate et al. (1985), did not describe the treatments well enough to permit categorization, although it seemed to evaluate a transitional model.

In Table 1, the elementary studies of Spanish-dominant children are listed according to these treatment categories, with the highest-quality studies listed first. That is, randomized studies are listed first, then matched multi-year studies, then matched one-year studies. The studies will be discussed in the same order.

Studies of Paired Bilingual Programs

Nine qualifying studies compared paired bilingual and English immersion programs. Plante (1976) randomly assigned 55 Spanish-dominant, Puerto Rican children in a New Haven, Connecticut, elementary school to a paired bilingual model or to English-only instruction. Two cohorts of experimental students were taught all of their

basic skills (reading, writing, math, science, social studies) in Spanish in first and second, or second and third grades. At different times of the day, they received English instruction designed to transition them to English-only instruction. After two years, second-graders in the program scored significantly higher on an English reading test than their English-only counterparts ($ES=+0.62$). Differences were not significant for third-graders, but still favored the bilingual group ($ES=+0.24$). Not surprisingly, the bilingual students also scored substantially higher on Spanish reading measures.

In a very similar study, Huzar (1973) randomly assigned two groups of Spanish-dominant, Puerto Rican children in Perth Amboy, New Jersey, to bilingual or English-only classes. One group ($N=81$) was in the study in first and second grades, and the other ($N=79$) was in the study in first through third grades. The experimental and control groups were well matched on IQ, SES, and initial achievement. As in the Plante (1976) study, students in the paired bilingual group had two teachers. One taught reading in Spanish for 45 minutes daily, while the other taught reading in English for the same amount of time. While the two groups did not differ after two years, children who were in the program for three years (Grades 1-3) scored higher than the control group in English reading. Using the control group standard deviation the effect size would be $+0.68$, but experimental and control standard deviations are very different. Using a pooled standard deviation yields a more conservative $ES = +0.31$.

The Huzar (1973) and Plante (1976) studies are particularly important, despite taking place more than a quarter century ago. Both are multi-year experiments that, due to use of random assignment, can rule out selection bias as an alternative explanation for the findings. Both started with children in the early elementary grades and followed them

for two to three years. Both used paired bilingual reading instruction by different teachers in Spanish and English, with transition to all-English instruction by second or third grade. The use of both Spanish and English reading instruction each day more resembles the experience of Spanish-dominant students in two-way bilingual programs (see Calderón & Minaya-Rowe, 2003) than it does transitional bilingual models, which delay English reading to second or third grade.

In the mid-1970s, the American Institutes of Research (AIR) produced a series of reports on bilingual programs around the U.S. (Campeau, Roberts, Oscar, Bowers, Austin, & Roberts, 1975). These are of some interest, with one major caveat: The AIR researchers were looking for *exemplary* bilingual programs. They began with 96 candidates and ultimately winnowed this list down to eight. Programs were excluded if data were unavailable, not because they failed to show positive effects of bilingual programs. Nevertheless, these sites were chosen on their reputations for excellence, and a site would clearly be less likely to submit data if the data were not supportive of bilingual education. Also, the Campeau et al. (1975) evaluations were organized as successive one-year studies, meaning that pretests after the first treatment year (K or 1) are of little value. For reasons described earlier, one-year evaluations of bilingual education are likely to be biased against the bilingual group on early English measures. With these cautions in mind, the Campeau et al. (1975) studies are described below.

A study in Corpus Christi, Texas, evaluated a paired bilingual program in three schools. The kindergarten program made extensive use of both Spanish and English instruction and reading materials in both languages, but the emphasis was on Spanish (90% of the instruction). A control group, consisting of students in three different

schools, was taught only in English. In the 1972-73 cohort, experimental and control classes were well matched on both English and Spanish measures.

At the end of kindergarten, the control group was slightly ahead on a standardized test of letters and sounds, but the bilingual group was slightly ahead on an English test of general ability. A second bilingual kindergarten cohort (1973-74) also slightly outscored the control group on general ability in English. The first-graders in 1973-74, who were the kindergarteners in the earlier analysis, ended the year with the bilingual students scoring 50% of a grade equivalent ahead of controls in SRA reading and substantially ahead of controls on general ability in English ($ES=+0.45$). They also were far ahead in Spanish ability.

A study in Houston also reported by Campeau et al. (1975) followed three cohorts of students in seven paired bilingual and two immersion schools. On a kindergarten pretest of English ability the students in the immersion groups scored substantially higher in all three cases, but at the end of kindergarten the bilingual classes were substantially higher on the English ability test. Controlling for pretests, these differences were highly significant ($p<.001$). Bilingual first graders in the second cohort and first and second graders in the third cohort (the former kindergartners) consistently outscored students who were in the immersion program.

Cohen (1975) compared two schools serving many Mexican Americans in Redwood City, California. One school was using what amounts to a two-way bilingual program, in that Spanish-dominant students and English-dominant students were taught in both Spanish and English. However, from the perspective of the ELLs, the treatment was the same as a paired bilingual model. Three successive cohorts were compared at the

two schools: grades K-1, 1-2, and 1-3. In each case, students were pretested and posttested on a broad range of English reading measures. In all cohorts, Mexican-American students were well matched on English and Spanish pretests. At posttest, there were no significant differences, adjusting for pretests. The data did not allow for computation of effect sizes.

A study in Corpus Christi, Texas, by J.R. Maldonado (1977) compared Mexican-American children in paired bilingual and English-only classes in grades 1-5, and found no differences at any grade, controlling for first-grade pretests. A study by Alvarez (1975) followed Mexican-American children in Austin, Texas, from first to second grades. There were no differences between children taught in English and those taught in English and Spanish.

Two of the studies carried out by Campeau et al. (1975) had one-year durations. A one-year study in Kingville, Texas, reported significantly greater gains on English SRA achievement tests for kindergartners in a paired bilingual program than for English immersion kindergartners in all six classroom pairs assessed.

Another one-year study in Santa Fe, New Mexico, compared paired bilingual and immersion programs for Spanish-dominant students. Pre- and posttests are reported for each year but only first grade is interpretable, as pretests for other years had already been affected by the treatments. Parents chose to place their children in bilingual or English programs, and apparently parents of higher-achieving children chose the bilingual group, as pretest scores were higher in that group. However, the bilingual group also gained more in English reading than the English-only group. No standard deviations were given, so effect sizes for pretest differences and gains could not be computed.

Studies of One-Year Transitional Bilingual Education

J.A. Maldonado (1994) carried out a small, randomized study involving English language learners who were in special education classes in Houston. Twenty second- and third-graders with learning disabilities were randomly assigned to one of two groups. A bilingual group was taught mostly in Spanish for a year, with a 45-minute ESL period. During a second year, half of the instruction was in English, half in Spanish. In a third year, instruction was only in English. The control group was taught in English all three years.

Children were pretested on the CTBS and then posttested on the CTBS three years later. At pretest, the control group scored nonsignificantly higher than the bilingual group, but at posttest the bilingual group scored far higher. Using the means and standard deviations presented in the article, the effect size would be +8.33, but using the given values of t , the effect size is +2.21, a more credible result.

A study in Alice, Texas, compared Spanish-dominant students in bilingual and English immersion programs starting in kindergarten, for a two-year study. The treatment involved teaching kindergartners in Spanish, and then transitioning them to English reading instruction in first grade. While kindergartners were comparable at pretest on English measures of general ability, bilingual students scored substantially higher on a Spanish ability test. At posttest (controlling for pretests), bilingual students scored substantially better in English reading at the end of first grade (after two years of bilingual education).

Study of Two-Year Transitional Bilingual Education

One of the most widely cited studies of bilingual education is a longitudinal study by Ramirez et al. (1991) that compared Spanish-dominant students in English immersion schools to two forms of bilingual education: early exit (transition to English in grades 2-4) and late-exit (transition to English in Grades 5-6). Schools in several districts were followed over four years. Immersion and early-exit students were well matched, but late-exit students were lower than their comparison groups in SES and their schools had much lower proportions of native English speakers. For these reasons, no direct comparisons were made by the authors between late-exit and other schools.

The comparison of early-exit transitional bilingual education and English immersion is the important contribution of the Ramirez et al. (1991) study. It involved four schools, each of which provided both programs. The children in the two programs were well matched on kindergarten pretests, socioeconomic status, preschool experience, and other factors. They were tested on the English CTBS each spring in grades 1-3. In English reading, the early-exit children scored significantly better than English immersion students at the end of first grade. By third grade, these differences were in the same direction but were not statistically significant, controlling for premeasures.

The Ramirez et al. study was so important in its time that the National Research Council convened a panel in 1991 to review it and a study by Burkheimer et al. (1989). The panel's report (Meyer & Fienberg, 1992) supported the conclusions of the Ramirez et al. comparison of the early-exit and immersion programs in grades K-1.

Meyer & Fienberg (1992) did not support the conclusions of the Burkheimer et al. study on effects of various bilingual and immersion models, due to lack of clear

comparisons of alternative treatments (among many other problems), and the Burkheimer et al. study was excluded from this review for similar reasons.

Study of Unspecified Bilingual Education

Saldate, Mishra, & Medina (1985) studied 62 children in an Arizona border town who attended immersion or bilingual schools. The bilingual treatment was unspecified, but appeared to be a transitional model like the one studied by Ramirez et al. (1991). The children were individually matched on the Peabody Picture Vocabulary Test in first grade. At the end of second grade, the bilingual students scored nonsignificantly lower on the English Metropolitan Achievement Test (MAT) ($ES=-0.29$) and higher on the Spanish MAT ($ES=+0.46$). This was to be expected, as they had not yet transitioned to English instruction. At third grade, however, the bilingual students (who had now transitioned to English-only instruction) substantially outperformed the immersion students both in English ($ES=+1.47$) and in Spanish ($ES=+6.40$). This study's small size means that its results should be interpreted cautiously, especially as the number of pairs dropped from 31 to 19 between second and third grades.

Studies Involving Languages Other Than Spanish

Three qualifying studies involved languages other than Spanish. These are reviewed separately not for this reason alone, but also because the languages (French in Louisiana, Cherokee, and Choctaw) are "heritage languages," whose use was intended as much to show respect to children's cultures as to help non-English speakers succeed in reading.

Morgan (1971) carried out a study of almost 200 children of French-speaking parents in rural Louisiana. Existing groups of first graders, assigned to bilingual or monolingual classes, were followed for a year. In the bilingual classes, children were taught in both French and English. The two groups were virtually identical on English tests of mental abilities and readiness at the beginning of first grade. At the end, the children taught in the bilingual classes scored higher on four English reading measures, with a median difference of +0.26. Differences were significant on measures of word reading and paragraph reading, but not vocabulary or word study skills. It is important to note, however, that the children in this study were probably English proficient. Their parents may have spoken French at home, but both experimental and control students scored well at pretest on an English mental abilities test.

Bacon, Kidd, & Seaberg (1982) evaluated a bilingual program for Cherokee students in Northeastern Oklahoma. The program introduced Cherokee language and reading materials to supplement English materials. This was clearly a heritage language approach; children apparently spoke English, and 28% of them did not speak Cherokee. The experimenters tested children as eighth graders. Two groups of children had attended the bilingual school in grades 1-5 or 2-5. Matched children from other schools taught only in English were the control group. The groups were not well matched, however; the control group had many more girls, and higher IQ's, father's education, and grade point averages. On the eighth-grade tests all groups were nearly identical, but after using regression analyses to control for matching factors, the bilingual groups scored higher. However, one of the control variables was grade point average, which was higher in the control group, so the analysis may have overadjusted the control scores.

A one-year study of 63 Choctaw second graders in Mississippi compared a bilingual program in Choctaw and English to English-only instruction (Doebler & Mardis, 1980-81). There were no significant differences on an English reading measure, controlling for pretests.

Studies of Secondary Reading

Two qualifying studies evaluated programs that introduced Spanish-language instruction to ELLs in the secondary grades. Both of these used random assignment.

Covey (1973) randomly assigned 200 low-achieving Mexican-American ninth graders to bilingual or English-only classes. The experimental intervention is not described in any detail, but it clearly involved extensive use of Spanish to supplement English in reading, English, and math. The groups' scores were nearly identical at pretest, but at posttest the bilingual students scored significantly better on the Stanford Diagnostic Reading Test ($ES=+0.82$).

Kaufman (1968) evaluated a program in which low-achieving Spanish-speaking seventh graders were randomly assigned to bilingual or English-only conditions in two New York junior high schools. One school participated in the program for a year and the other for two years. In the bilingual classes, students received three or four periods of Spanish reading instruction each week, while controls were in art, music, or health education. On standardized tests of word and paragraph meaning, there were no significant differences in the two-year school, but in the one-year school significant differences favored the bilingual group on one of two word meaning tests. Results of

paragraph meaning tests favored the bilingual group, though not significantly. The data presented did not permit computation of effect sizes.

The secondary studies point to the possibility that providing native language instruction to low-achieving ELLs in secondary school may help them with English reading. This application is worthy of additional research (also see Klingner & Vaughn, 2004).

Canadian Studies of French Immersion

There are several Canadian studies (e.g., Lambert & Tucker, 1972; Genesee & Lambert, 1983; Day & Shapson, 1988; Barik & Swain, 1978) that have played an important role in debates about bilingual education. These are studies of French immersion programs, in which English speaking children are taught entirely or primarily in French in the early elementary years. Rossell & Baker (1996) emphasized these studies as examples of “structured English immersion,” the approach favored in their review. However, Willig (1985) and other reviewers have excluded them. These studies do not meet the inclusion standards of this review because the Anglophone children are learning a useful second language, not the language for which they will be held accountable in their later schooling. Although many of the studies took place in Montreal, the children lived in English-speaking neighborhoods, and attended schools in an English system. The focus of this review is on bilingual education used to help children succeed in the language in which they will be taught in the later grades, but the French immersion children in Canada are headed to English secondary schools. Further, these studies all involve voluntary programs, in which parents wanted their children to

learn French, and the children in these studies were generally upper middle class, not disadvantaged.

Because French immersion programs were voluntary, children who did not thrive in them could be and were routinely returned to English-only instruction. This means that the children who complete French immersion programs in Canada are self-selected, relatively high achievers. Most importantly, the “bilingual” programs to which French immersion is compared are nothing like bilingual education in the U.S. At most, children receive 30 to 40 minutes daily of French as a second language, with far less time in French reading instruction than a U.S. student in a bilingual program would receive in English during and after transition (see Genesee & Lambert, 1983). Yet in many studies, English comparison groups were not learning French at all. In the widely cited study by Lambert & Tucker (1972), Anglophones in French immersion classes were compared to Anglophones taught only in English, and to francophones taught only in French. Ironically, studies of this kind, cited by Rossell & Baker (1996) as comparisons of immersion and bilingual education, are in fact comparisons of immersion and *monolingual* education. If they existed, Canadian studies of, say, Spanish speakers learning French in francophone schools in Quebec or English in Anglophone schools in the rest of Canada would be relevant to this review, but studies of voluntary immersion programs as a means to acquiring French as a second language are only tangentially relevant.

While the Canadian immersion studies are not directly relevant to the question of the effectiveness of bilingual programs for ELLs learning the societal language, they are nevertheless interesting in gaining a broader understanding of the role of native language

in foreign language instruction. As a group, these studies are of high methodological quality. Quite in contrast to U.S. studies, however, the focus of the Canadian studies is on whether or not French immersion harms the English language development of native English speakers. It is taken as obvious that French all day will produce more facility in French than 30 to 40 minutes daily in second language classes.

Lambert & Tucker (1972) carried out the foundational study of French immersion in Canada. It compared Anglophone children taught completely in French from kindergarten and first grade, with some English instruction in Grades 2-4, to matched Anglophone children taught in English and to francophone children taught in French. At the end of first grade, immersion children scored far below children taught in English on English reading measures. And, while their spoken French was much worse than the French controls, their French reading was as good as that of the native speakers. A study of a second cohort found similar results at first grade. At second grade, however, the immersion students had almost caught up to the English-only students, and there were no differences in third or fourth grade in either English (compared to English-only Anglophones) or French (compared to French-only francophones). A followup to grades 5-6 found the same patterns (Bruck, Lambert, & Tucker, 1977)

The finding of no differences was taken as a vindication of French immersion, as the Anglophone children suffered no loss in English reading and gained fluent reading and speaking skills in an important second language. Because the comparison students were taught in only one language, however, there is no “bilingual” group to which immersion could be compared.

Other French immersion studies followed a similar paradigm. For example, Barik & Swain (1975) studied a program in Ottawa, which also found similar second grade English reading performance for Anglophone children taught entirely in French in Grades K-1, with 60 minutes daily of English instruction in second grade, compared to Anglophone children taught only in English. There were no differences in English reading by the end of second grade. Another Ontario study by Barik, Swain, & Nwanunobi (1977) compared a “partial French immersion” program (essentially, a paired bilingual program with 50% of instruction in each language) to English-only instruction. The English-only students performed better in English reading through third grade, but in grades 4 and 5 the two groups were similar, and the partial immersion students were fluent in French.

Overall, the Canadian studies paint a consistent picture. At least for the overwhelmingly middle-class students involved, French immersion had no negative effect on English reading achievement, and it gave students facility in a second language. The relevance to the U.S. situation is in suggesting that similar second-language immersion programs, as well as two-way bilingual programs for English proficient children, are not likely to harm English reading development. However, the relevance of these studies to any context in which the children of immigrants are expected to learn the language that will constitute success in their school and in the larger society is unclear.

Comparisons of Paired Bilingual and Transitional Bilingual Programs

As noted earlier, many of the programs with the strongest positive effects for English language learners used a paired bilingual approach, in which children were taught reading in both English and their native language at different times each day from the beginning of their schooling. This approach contrasts with transitional bilingual education (TBE) models in which children are first taught to read primarily in their native language, and only then transitioned gradually to English-only instruction. Two studies have compared reading outcomes of these two bilingual approaches.

A longitudinal study by Gersten & Woodward (1995) initially favored paired bilingual instruction over TBE, but later found them to be equivalent. This study was carried out with Spanish-dominant ELLs in 10 El Paso elementary schools. Five schools used a program in which all subjects were taught in English, but Spanish instruction was also provided, for 90 minutes daily in first grade declining to 30 minutes a day in fourth grade. The transitional bilingual program involved mostly Spanish instruction with one hour per day for ESL instruction, with gradual transition to English completed in the fourth or fifth grade. The children were well matched demographically on entry to first grade, and scored near zero on a measure of English language proficiency. In grades 4, 5, 6, and 7, Iowa Tests of Basic Skills were compared for the two groups. On Total Reading, the paired bilingual students scored significantly higher than the transitional bilingual students in fourth grade ($ES=+0.31$), but the effects diminished in fifth grade ($ES=+0.18$), and were very small in sixth ($ES=+0.06$) and seventh grades ($ES=+0.08$). Tests of language and vocabulary showed similar patterns. This pattern is probably due to the fact that the transitional bilingual students had not completed their transition to

English in fourth and fifth grades. When they had done so, by sixth grade, their reading performance was nearly identical.

A one-year study of Spanish-dominant kindergartners by Pena-Hughes & Solis (1980) could not be located, but information provided by Willig (1985) indicated that it found a strong advantage of a paired bilingual approach over a TBE model. As the paired bilingual program provided more time in English instruction, however, a longer study would be needed to establish the relative effects.

Research comparing alternative bilingual models is far from conclusive, but nothing suggests that it is harmful to children's reading performance to introduce both native language and English reading instruction at different times each day.

Discussion

The most important conclusion from research on language of instruction is that there are far too few high-quality studies of this question. Willig (1985) and Rossell & Baker (1996) agree on very little, but both of these reviews call for randomized, longitudinal evaluations to produce a satisfying answer to this critical question. Of course, many would argue that randomized evaluations are needed on most important questions of educational practice (see, for example, Mosteller & Boruch, 2002; Slavin, 2003), but in bilingual education, this is especially crucial due to the many inherent problems of selection bias in this field. Further, this is an area in which longitudinal, multi-year studies are virtually mandatory, to track children initially taught in their native language through their transition to English. Finally, while randomized, longitudinal

studies of this topic are sorely needed, there are simply too few experimental studies of all kinds, including ones with matched experimental and control groups.

With these concerns in mind, however, research on language of instruction does yield some important lessons at least worthy of further study. Across 18 qualifying studies of all types of programs, 13 found effects favoring bilingual education and 5 found no differences. None of the studies found results favoring English immersion.

The largest group of studies focused on elementary reading instruction for Spanish-dominant students. Nine of 13 studies in this category favored bilingual approaches, and four found no differences. Eight of the studies provided sufficient information for computation of effect sizes. Among these, the median effect size was +0.52. Including secondary studies and heritage language studies, the median effect size from a total of 12 studies is also +0.52. This effect size is higher than the estimate of +0.21 given by Greene (1997), but Greene did not locate the Campeau et al. (1975) studies that added several positive effect sizes. The median effect size calculated in this review is closer to Greene's effect size estimate of +0.43 for studies using random assignment. However, the findings of this review strongly conflict with the conclusions of Rossell & Baker (1996).

It was surprising to find that most of the methodologically adequate studies located evaluated forms of bilingual education quite different from those commonly used in recent years. These are paired bilingual programs, in which children are taught to read in English and in their native language at different times each day from the beginning of their time in school. Another category of programs provided just one year of native-language instruction before transition to English-only reading. Paired bilingual strategies

were used in two of the randomized studies (Huzar, 1973; Plante, 1976), and in a study of a one-year transitional program (J. A. Maldonado, 1994). These practices contrast sharply with practices in transitional bilingual education, in which children are typically taught to read in their native language from kindergarten to grades two, three, or four, and then transitioned to reading. Only one qualifying study, the classic Ramirez et al. (1991) experiment, clearly evaluated a transitional bilingual program. Ironically, the treatment evaluated by Ramirez et al. was called “early exit,” even though its “exit” year, second grade at the earliest, was later than that in any other qualifying study of beginning reading.

There are several reasons that paired bilingual interventions may be so prevalent among the studies reviewed. First, most of the studies reviewed took place in the 1970’s, when Title VII was new. At that time, paired bilingual models were popular. Second, for reasons discussed earlier in this review, studies of transitional bilingual education are very difficult to do, as they should begin in kindergarten and continue past the point of transition. A four-year longitudinal study (like the Ramirez et al. study) would be required to follow children from kindergarten to third grade. Allowing for student mobility, such a study must start with a large sample in order to end up with sufficient numbers of students. The U. S. Department of Education has recently funded two matched and one randomized longitudinal study to evaluate transitional bilingual education, but before these only the Ramirez et al. (1991) study had the resources to carry out an investigation of this kind.

It is important to note that most of the studies that did not qualify for inclusion also used paired bilingual modes, not transitional bilingual models. A key exception was

a series of studies by Thomas & Collier (2002) that followed children who had been in transitional programs but lacked pretest measures from before the TBE interventions began.

Because of the dearth of studies of TBE, it is not currently possible to say with confidence whether paired bilingual models are more effective than transitional models. However, two studies have made this comparison. One, by Gersten & Woodward (1995), found differences favoring paired bilingual strategies in Grades 4 and 5, but not in Grades 6 and 7. Another, by Pena-Hughes & Solis (1980), found a strong advantage for a paired bilingual strategy. Given these findings, and more importantly, the strong support for paired bilingual methods seen in this review, it is worthwhile to speculate about why paired methods might be beneficial.

Teaching a Spanish-speaking English language learner in Spanish can be expected to establish the alphabetic principle, the idea that words are composed of distinct sounds represented by letters (see National Reading Panel, 2000). Early in their reading instruction, children learn to combine letters and sounds into words they know. This process is very difficult if children must form letters and sounds into words they don't know, so it may greatly facilitate phonetic development to learn the alphabetic principle in a familiar language rather than an unfamiliar one. Once a Spanish-speaking child can confidently decode Spanish text, he or she should be able to make an easy transfer to decoding any alphabetic language, such as English, by learning a modest number of new sounds for particular graphemes (Lindsey, Manis, & Bailey, 2003). Several of the studies of paired bilingual instruction clearly described a process of teaching Spanish reading phonetically and then planfully transferring those skills to English decoding.

Rather than confusing children, as some have feared, reading instruction in a familiar language may serve as a bridge to success in English, as phonemic awareness, decoding, sound blending, and generic comprehension strategies clearly transfer among languages that use phonetic orthographies, such as Spanish, French, and English (see August, 2002; August, Calderón, & Carlo, 2001; August & Hakuta, 1997; Durgonolglu, Nagy, & Hancin-Bhatt, 1993; Fitzgerald, 1995; Garcia, 2000; Lee & Schallert, 1997; Lindsey, Manis, & Bailey, 2003).

Only two studies of secondary programs met the inclusion criteria, but both of these were very high quality randomized experiments. Covey (1973) found substantial positive effects of Spanish instruction for low-achieving ninth graders, while Kaufman (1968) found mixed, but slightly positive, effects of a similar approach with low-achieving seventh graders.

As noted previously, research on language of instruction may suffer from publication bias, the tendency for journals to publish only articles that find significant differences. However, dissertations and technical reports (e.g., Covey, 1973; Huzar, 1973; Plante, 1976) less likely to suffer from publication bias also tended to favor bilingual programs.

Teaching reading in two languages, with appropriate adaptations of the English program for the needs of English language learners, may represent a satisfactory resolution to the acrimonious debates about bilingual education. Proponents of bilingual education want to launch English language learners with success while maintaining and valuing the language they speak at home. Opponents are concerned not so much about the use of native language, but about delaying the use of English. Paired bilingual

models immerse children in both English reading and native language reading at the same time. They are essentially half of a two-way bilingual model; by encouraging English proficient students to also take Spanish reading, any school with a paired bilingual model can readily become a two-way program, offering English-only children a path to early acquisition of a valuable second language (see Calderón & Minaya-Rowe, 2003; Howard et al., 2003).

Language of instruction must be seen as only one aspect, however, of instructional programming for English language learners. As many previous reviewers have concluded, quality of instruction is at least as important as language of instruction. (For reviews of effective programs and practices for ELLs, see August & Hakuta, 1997; Fitzgerald, 1995; Klingner & Vaughn, 2004, Slavin & Cheung, 2003.)

Clearly, there is much more we need to know about the role of native language instruction in reading. The research reviewed in this article may represent the best experimental studies currently available, but better evidence is needed. Longitudinal experiments using random assignment of students to alternative treatments are particularly needed. Both qualitative and quantitative research are needed to illuminate the conditions under which native language instruction may be beneficial for developing English reading skills, and to explain these effects. Research systematically varying program components and research combining quantitative and qualitative methods are needed to more fully understand how various interventions affect the development of reading skills among English language learners. It is time to end the ideological debates, and to instead focus on good science, good practice, and sensible policies for children

whose success in school means so much to themselves, their families, and our nation's future.

References

- Alvarez, J. (1975). *Comparison of academic aspirations and achievement in bilingual Versus Monolingual Classrooms*. Ph.D. dissertation, University of Texas at Austin.
- August, D., Calderón, M., & Carlo, M. (2001). The transfer of skills from Spanish to English: A study of young learners. *NABE News*, May/June, 11-12, 42.
- August, D. (2002). *English as a second language instruction: Best practices to support the development of literacy for English language learners*. Baltimore: Johns Hopkins University, Center for Research on the Education of Students Placed at Risk.
- August, D., & Hakuta, K. (1997). *Improving schooling for language-minority children: A research agenda*. Washington, DC: National Research Council.
- Bacon, H., Kidd, G., & Seaberg, J. (1982). The effectiveness of bilingual instruction with Cherokee Indian students. *Journal of American Indian Education*, 21 (2), 34-43.
- Baker, K. (1987). A meta-analysis of selected studies in the effectiveness of bilingual education. *Review of Educational Research*, 57, 351-362.

Baker, K., & de Kanter, A. (1981). *Effectiveness of bilingual education: A review of the literature*. (Final draft report). Washington, DC: Office of Technical and Analytic Systems, U.S. Department of Education.

Baker, K., & de Kanter, A. (1983). An answer from research on bilingual education. *American Education*, 56, 157-169.

Balasubramonian, K., Seelye, H., and Elizondo de Weffer, R. (1973). *Do bilingual education programs inhibit English language achievement: A report on an Illinois experiment*. Paper presented at the 7th Annual Convention of Teachers of English to Speakers of Other Languages, San Juan.

Barik, H. and Swain, M. (1975). Three year evaluation of a large-scale early grade French immersion program: The Ottawa study. *Language Learning*. Vol. 25. No. 1. pp. 1-30.

Barik, H. and Swain, M. (1978). *Evaluation of a bilingual education program in Canada: The Elgin study through grade six*. Commission Interuniversitaire Suisse de Linguistique Appliquee. Switzerland.

Barik, H., Swain, M. and Nwanunobi, E. A. (1977). English-French bilingual education: The Elgin study through grade five. *Canadian Modern Language Review*. 33, 459-475.

Bates, E. & May B. (1970). *The effects of one experimental bilingual program on verbal ability and vocabulary of first grade pupils*. Ph.D. dissertation, Texas Tech University.

Bruck, M, Lambert, W. E., and Tucker, G. R. (1977). Cognitive Consequences of Bilingual Schooling: The St. Lambert Project Through Grade Six. *Linguistics*. (24), 13-33.

Burkheimer, G. J., Conger, A.J., Dunteman, G.H., Elliott, B.G., and Mowbray, K.A. (1989). *Effectiveness of services for language-minority limited-English-proficient students*. Washington, DC: U.S. Department of Education.

Calderón, M. (2001). Curricula and methodologies used to teach Spanish-speaking limited English proficient students to read English. In R.E. Slavin & M. Calderón (Eds.), *Effective Programs for Latino Students*. Mahwah, NJ: Erlbaum.

Calderón, M., & Minaya-Rowe, L. (2003). *Designing and implementing two-way bilingual programs*. Thousand Oaks, CA: Corwin.

Carlisle, J.F. & Beeman, F.F. (2000). The effects of language of instruction on the reading and writing achievement of first-grade Hispanic children. *Scientific Studies of Reading*, 4 (4), 331-353.

Campeau, P. L., Roberts, A. Oscar H., Bowers, John E., Austin, M., and Roberts, S. J.

(1975). *The identification and description of exemplary bilingual education programs*. Palo Alto, CA: American Institutes for Research.

Carlo, M.S., August, D., McLaughlin, B., Snow, C.E., Dressler, C., Lippman, D., Lively,

T., & White, C. (2004). Closing the gap: Addressing the vocabulary needs of English language learners in bilingual and mainstream classrooms. *Reading Research Quarterly*, 39 (2), 188-215.

Cohen, A. D. (1975). *A sociolinguistic approach to bilingual education*. Rowley, MA:

Newbury House Press.

Cooper, H. (1998). *Synthesizing research (3rd ed.)*. Thousand Oaks, CA: Sage.

Cooper, H.M., & Hedges, L.V. (Eds.) (1994). *The handbook of research synthesis*. New

York: Russell Sage Foundation.

Covey, D. D. (1973). *An analytical study of secondary freshmen bilingual education and*

its effects on academic achievement and attitudes of Mexican American students.

Ph.D. dissertation, Arizona State University.

Curiel, H., Stenning, W., and Cooper-Stenning, P. (1980). Achieved ready level, self-esteem, and grades as related to length of exposure to bilingual education.

Hispanic Journal of Behavioral Sciences, Vol. 2, pp. 389-400.

Danoff, M. N., Arias, B.M., Coles, G. J., and Others. (1977a). *Evaluation of the Impact of ESEA Title VII Spanish/English Bilingual Education Program*. Palo Alto, CA:

American Institutes for Research.

Danoff, M.N., Coles, G.J., McLaughlin, D.H. & Reynolds, D.J. (1978). *Evaluation of the Impact of ESEA Title VII Spanish/English Bilingual Education Programs, Vol.*

III: Year Two Impact Designs. Palo Alto, CA: American Institutes for Research.

Danoff, M. N., Coles, G. J., McLaughlin, D. H., and Reynolds, D. J. (1977b). *Evaluation of the Impact of ESEA Title VII Spanish/English Bilingual Education Programs,*

Vol. I: Study Design and Interim Findings. Palo Alto, CA: American Institutes for Research.

Day, E. M., & Shapson, S. M. (1988). *Provincial assessment of early and late French immersion programs in British Columbia, Canada*. Paper presented at the April

meetings of the American Educational Research Associates. New Orleans.

de Weffer, R. (1972). *Effects of first language instruction in academic and psychological development of bilingual children*. Ph.D. dissertation, Illinois Institute of Technology.

Doebler, L.K. & Mardis, L.J. (1980-81). Effects of a bilingual education program for Native American children. *NABE Journal*, 5 (2), 23-28.

Durgunoglu, A., Nagy, W. E., & Hancin-Bhatt, B. J. (1993). Cross-language transfer of phonological awareness. *Journal of Educational Psychology*, 85 (3), 453-465.

El Paso Independent School District. (1987). *Interim report of the five-year bilingual education pilot 1986-1987 school year*. Office for Research and Evaluation. El Paso.

El Paso Independent School District. (1990). *Bilingual education evaluation: The sixth year in a longitudinal study*. El Paso, TX: Office for Research and Evaluation.

El Paso Independent School District. (1992). *Bilingual education evaluation*. El Paso, TX: Office for Research and Evaluation.

Fitzgerald, J. (1995). English as a second language instruction in the United State: A research review. *Journal of Reading Behavior*, 27, 115-152.

- Garcia, G. (2000). Bilingual children's reading. In M.L. Kamil, P.B. Mosenthal, P.D. Pearson, & R. Barr (Eds.), *Handbook of Reading Research*, Vol. III (pp. 813-834). Mahway, NJ: Erlbaum.
- Genesee, F., Lambert, W. E., and Tucker, G. R. (1977). *An experiment in trilingual education*. Montreal, Canada: McGill University.
- Genesee, F. & Lambert, W. E. (1983). Trilingual education for majority-language children. *Child Development*, 54. 105-114.
- Gersten, R. (1985). Structured immersion for language minority students: Results of a longitudinal evaluation. *Educational Evaluation and Policy Analysis*, 7 (3), 187-196.
- Gersten, R., & Woodward, J. (1995). A longitudinal study of transitional and immersion bilingual education programs in one district. *The Elementary School Journal*, 95 (3), 223-239.
- Greene, J.P. (1997). A meta-analysis of the Rossell & Baker review of bilingual education research. *Bilingual Research Journal*, 21 (2/3).
- Grigg, W., Daane, M., Jin, Y., Campbell, J. (2003). *The nations's report card: Reading 2002*. Washington, DC: US Department of Education.

Hakuta, K., Butler, Y.G., & Witt, D. (2000). *How long does it take English learners to attain proficiency?* The University of California Linguistic Minority Research Institute, Policy Report 2000-1.

Howard, E.R., Sugarman, J., & Christian, D. (2003). *Two-way immersion education: What we know and what we need to know.* Baltimore, MD: Johns Hopkins University, Center for Research on the Education of Students Placed at Risk.

Huzar, H. (1973). *The effects of an English-Spanish primary grade reading program on second and third grade students.* M.Ed. thesis, Rutgers University.

Kaufman, M. (1968). Will instruction in reading Spanish affect ability in reading English?" *Journal of Reading*, 11, 521-527.

Klingner, J. K., & Vaughn, S. (2004). Specific strategies for struggling second language readers and writers. In T. L. Jetton & J. A. Dole (Eds.), *Adolescent Literacy Research and Practice*. New York: Guilford.

Lambert, W. E., and Tucker, G. R. (1972). *Bilingual education of children: The St. Lambert Experience*. Rowley, Quebec: Newbury House.

- Layden, R. G. (1972). *The relationship between the language of instruction and the development of self-concept, classroom climate, and achievement of Spanish speaking Puerto Rican children*. Ph.D. dissertation, University of Maryland.
- Lee, J., & Schallert, D. L. (1997). The relative contribution of L2 language proficiency and L1 reading ability to L2 reading performance: A test of the threshold hypothesis in an EFL context. *TESOL Quarterly*, 31, 713-739.
- Legarreta, D. (1979). The effects of program models on language acquisition by Spanish-speaking children. *TESOL Quarterly*, 13, (4), 521-534.
- Lindsey, K. A., Manis, F. R., & Bailey, C. E. (2003). Prediction of first-grade reading in Spanish-speaking English-language learners. *Journal of Educational Psychology*, 95 (3), 482-494.
- Lum, J. B. (1971). *An effectiveness study of English as a second language (ESL) and Chinese bilingual methods*. Ph.D. dissertation, University of California at Berkeley.
- Maldonado, J.A. (1994). Bilingual special education: Specific learning disabilities in language and reading. *Journal of Education Issues of Language Minority Students*, 14, 127-147.

- Maldonado, J.R. (1977). *The effect of the ESEA Title VII program on the cognitive development of Mexican American students*. Unpublished doctoral dissertation, University of Houston, Houston, TX.
- Matthews, T. (1979). *An investigation of the effects of background characteristics and special language services on the reading achievement and English fluency of bilingual students*. Seattle, WA: Seattle Public Schools, Department of Planning, Research, and Evaluation.
- Melendez, W. A. (1980). *The effect of the language of instruction on the reading achievement of limited English speakers in secondary schools*. Ph.D. dissertation. Loyola University of Chicago.
- Meyer, M.M. & Fienberg, S.E. (1992). *Assessing evaluation studies: The case of bilingual education strategies*. Washington, DC: National Academy of Sciences.
- Morgan, J.C. (1971). *The effects of bilingual instruction of the English language arts achievement of first grade children*. Ph.D. dissertation, Northwestern State University of Louisiana.
- Mosteller, F., & Boruch, R. (2002). *Evidence matters: Randomized trials in education research*. Washington, DC: Brookings Institution.

- National Center for Education Statistics (2004). *Language minorities and their educational and labor market indicators-recent trends*. Washington, DC: U.S. Department of Education.
- National Reading Panel (2000). *Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction*. Rockville, MD: National Institute of Child Health and Human Development.
- Olesini, J. (1971). *The effect of bilingual instruction on the achievement of elementary pupils*. Unpublished doctoral dissertation, East Texas State University, Commerce, TX.
- Pena-Hughes, E, and Solis, J. (1980). *ABC's*. McAllen, TX: McAllen Independent School District.
- Plante, A. J. (1976). *A study of effectiveness of the Connecticut "Pairing" model of bilingual/bicultural education*. Hamden, CT: Connecticut Staff Development Cooperative.
- Powers, S. (1978). *The influence of bilingual instruction on academic achievement and self-esteem of selected Mexican American junior high school students*. Ph.D. dissertation. University of Arizona.

Ramirez, J., Pasta, D.J, Yuen, S., Billings, D. K., and Ramey, D. R. (1991). *Final report: Longitudinal study of structural immersion strategy, early-exit, and late-exit transitional bilingual education programs for language-minority children*. San Mateo, CA: Aguirre International (Report to the U.S. Department of Education).

Ramos, M., Aguilar, J.V., and Sibayan, B.F. (1967). *The determination and implementation of language policy*. Quezon City, The Phillipines: Phillipine Center for Language Study: Monograph Series 2.

Reese, L., Garnier, H., Gallimore, R., & Goldenberg, C. (2000). Longitudinal analysis of the antecedents of emergent Spanish literacy and middle-school English reading achievement of Spanish-speaking students. *American Educational Research Journal*, 37 (3), 633-662.

Rosier, P & Holm, W. (1980). *The Rock Point Experience: A longitudinal study of a Navajo school program*. Washington DC: Center for Applied Linguistics.

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

Rossell, C.H. & Baker, K. (1996). The educational effectiveness of bilingual education.

Research in the Teaching of English, 30 (1), pp. 7-69.

Rossell, C., & Ross, J. (1986). The social science evidence on bilingual education.

Journal of Law and Education, 15, 385-419.

Rothfarb, S. H., Ariza, M.J. and Urrutia, R. (1987). *Evaluation of the Bilingual*

Curriculum Content (BCC) Project: A three-year study, final report. Miami:

Office of Educational Accountability, Dade County Public Schools.

Saldate, M., Mishra, S. P., & Medina, M. (1985). Bilingual instruction and academic

achievement: A longitudinal study. *Journal of Instructional Psychology, 12* (1),

24-30.

Skoczylas, R. V. (1972). *An evaluation of some cognitive and affective aspects of a*

Spanish bilingual education program. Ph.D. dissertation, University of New

Mexico.

Slavin, R.E. (1986). Best-evidence synthesis: An alternative to meta-analytic and

traditional reviews. *Educational Researcher, 15* (9), 5-11.

Slavin, R.E. (2003). Evidence-based education policies: Transforming educational

practice and research. *Educational Researcher, 31* (7), 15-21.

- Slavin, R.E., & Cheung, A. (2003). *Effective programs for English language learners: A best-evidence synthesis*. Baltimore, MD: Johns Hopkins University, Center for Research on the Education of Students Placed at Risk.
- Stern, C. (1975). *Final report of the Compton Unified School District's Title VII bilingual-bicultural project: September 1969 through June 1975*. Compton City, CA: Compton City Schools.
- Thomas, W.P., & Collier, V.P. (2002). *A national study of school effectiveness for language minority students' long-term academic achievement*. Santa Cruz, CA: University of California at Santa Cruz, Center for Research on Education, Diversity, and Excellence.
- Valladolid, L. A. (1991). *The effects of bilingual education of students' academic achievement as they progress through a bilingual program*. Ph.D. dissertation, United States International University.
- Willig, A. (1985). A meta-analysis of selected studies on the effectiveness of bilingual education. *Review of Educational Research*, 55 (3), 269-317.
- Willig, A. (1987). Examining bilingual education research through meta-analysis and narrative review: A response to Baker. *Review of Educational Research*, 57 (3).

- Wong-Fillmore, L., & Valadez, C. (1986). Teaching bilingual learners. In M.C. Wittrock (Ed.), *Handbook of research on teaching* (3rd Ed.). New York: Macmillan.
- Yap, K. O., Enoki, D. Y., and Ishitani, P. (1988, April). *LEP student achievement: Some pertinent variables and policy implications*. Paper presented at the annual meetings of the American Educational Research Association. New Orleans.
- Yeung, A.E., Marsh, H.W., & Suliman, R. (2000). Can two tongues live in harmony?: Analysis of the National Education Longitudinal Study of 1988 (NELS88) longitudinal data on the maintenance of home language. *American Educational Research Journal*, 37 (4), 1001-1026.